



UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS

FACULTAD DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

E.A.P. DE INGENIERÍA DE SISTEMAS

**Implementación de un datamart como una solución de
inteligencia de negocios para el área de logística de T-
Impulso**

Tesina

Para optar el Título de Ingeniero de Sistemas

AUTORES

Luis Palomino Paniora

Julio Yalán Castillo

LIMA – PERÚ
2013

FICHA CATALOGRÁFICA

APELLIDOS Y NOMBRES DE LOS AUTORES.

PALOMINO PANIORA, Luis

YALAN CASTILLO, Julio

TITULO DE LA TESINA: IMPLEMENTACIÓN DE UN DATAMART
COMO UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS PARA
EL ÁREA DE LOGÍSTICA DE T-IMPULSO

Bases de Datos/Data Warehouse e Inteligencia de negocios
(Lima, Perú 2013)

Tesina, Facultad de Ingeniería de Sistemas e Informática, Pregrado,
Universidad Nacional Mayor de San Marcos

Formato 28 x 20 cm

Páginas 91

DEDICATORIA:

La presente tesina está dedicada al esfuerzo de nuestras familias por brindarnos una buena educación y confiar en nosotros a lo largo de nuestra carrera.

AGRADECIMIENTOS

Agradecemos a los profesores de los cursos de Taller de tesis I y Taller de tesis II por sentarnos la base para realizar una buena tesina.

A nuestro asesor, Augusto Parcemón Cortez Vásquez por apoyarnos con sus consejos y sugerencias en el desarrollo de la presente tesina.

A la empresa T-IMPULSO por toda su ayuda y colaboración.

UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS

FACULTAD DE INGENIERIA DE SISTEMAS E INFORMÁTICA

ESCUELA ACADEMICO PROFESIONAL DE INGENIERÍA DE SISTEMAS

**IMPLEMENTACIÓN DE UN DATAMART
COMO UNA SOLUCIÓN DE INTELIGENCIA DE NEGOCIOS PARA
EL AREA DE LOGISTICA DE T-IMPULSO**

Autores: YALAN CASTILLO, JULIO y PALOMINO PANIORA, LUIS

Asesor: CORTEZ VASQUEZ, AUGUSTO

Título: Tesina, para optar el Título Profesional de Ingeniero de Sistemas

Fecha: Febrero del 2013

RESUMEN

Con la informatización y el crecimiento de las empresas, se ha elevado en gran medida la capacidad de generación y almacenamiento de datos, son tan grandes los volúmenes de datos que su análisis no puede realizarse con los métodos tradicionales existentes. Mientras mayor es la capacidad para almacenar datos, mayor es la incapacidad para extraer información realmente útil, y gran parte de la información importante para la toma de decisiones queda oculta y muchas veces no tomadas en cuenta debido a que no se pudo obtener esa valiosa información en el momento requerido. Además los sistemas transaccionales que se utilizan comúnmente en las organizaciones no son los adecuados para análisis y extracción de información histórica y relevante. El presente trabajo se orienta a implementar un DataMart, como una herramienta que permitirá desarrollar Inteligencia de Negocios sobre el área de Logística de una Empresa que brinda servicios de Tercerización y Outsourcing; con el presente proyecto de tesina se plantea la simplificación de los procedimientos de extracción, transformación y procesamiento de datos, para la obtención de información del comportamiento de los proveedores, y con esto brindar apoyo en la toma de decisiones al área de Logística en estudio.

Palabra claves: DataMart, Data Warehouse, Inteligencia de Negocios, Logística, Compras, Proveedores

**MAJOR NATIONAL UNIVERSITY OF DE SAN MARCOS
FACULTY OF SYSTEMS ENGINEERING AND INFORMATICS**

ACADEMIC PROFESSIONAL SCHOOL SYSTEMS ENGINEERING

**IMPLEMENTATION OF A DATAMART AS SOLUTION BUSINESS INTELLIGENCE
FOR LOGISTICS AREA T-IMPULSO**

Authors: YALAN CASTILLO, JULIO and PALOMINO PANIORA, LUIS

Adviser: CORTEZ VASQUEZ, AUGUSTO

Title: Tesina, para optar el Título Profesional de Ingeniero de Sistemas

Date: February del 2013

ABSTRACT

With computerization and the growth of companies has greatly increased generating capacity and data storage, are so large volumes of data analysis can't be done with existing traditional methods. The greater the ability to store data, the greater is the inability to extract truly useful information, and much of the important information for decision making is hidden, and often not taken into account because they could not get that valuable information when required, in addition to transactional systems that are commonly used in organizations are not adequate for analysis and extraction of relevant historical information. This work aims to implement a DataMart, as a tool that will develop Business Intelligence on the Logistic of a company that provides outsourcing services; with this thesis project raises the simplification of procurement procedures, processing and data processing, to obtain information from customer behavior, and thereby support decision-making in the logistic area under study.

Key words: DataMart, Data Warehouse, Business Intelligence, Logistic, shopping.

INDICE

DEDICATORIA:.....	4
AGRADECIMIENTOS.....	5
RESUMEN	6
ABSTRACT.....	7
INDICE.....	8
LISTA DE FIGURAS.....	11
LISTA DE TABLAS.....	12
INTRODUCCION	13
CAPITULO I. PLANTEAMIENTO METODOLÓGICO	16
1.1. ANTECEDENTES.....	17
1.2. DEFINICION DEL PROBLEMA	18
1.3. FORMULACION DEL PROBLEMA.....	18
1.3.1. PROBLEMA GENERAL	18
1.3.2. PROBLEMA ESPECIFICO 1	19
1.4. OBJETIVOS.....	19
1.4.1. OBJETIVO GENERAL	19
1.4.2. OBJETIVO ESPECIFICO 1	19
1.5. JUSTIFICACION	19
1.5.1. POBLACIÓN A LA QUE BENEFICIA LA INVESTIGACIÓN	19
1.5.2. JUSTIFICACIÓN DEL PROYECTO	20
1.5.3. FACTIBILIDAD DEL ESTUDIO.....	20
1.5.4. IMPACTO ESPERADO DE LOS RESULTADOS.....	21
1.5.5. OPORTUNIDAD PARA TRATAR EL PROBLEMA	21
1.6. DELIMITACION DE LA INVESTIGACION	22
1.6.1. DELIMITACIÓN DEL PROBLEMA	22
1.6.2. ASPECTOS DE ESTUDIO QUE COMPRENDE EL PROBLEMA	22
1.7. TIPO DE INVESTIGACIÓN AL QUE PERTENECE EL ESTUDIO	22
1.7.1. UNIDAD DE ANÁLISIS	22
1.7.2. JUSTIFICACIÓN DEL TIPO DE INVESTIGACIÓN QUE SE PROPONE	22
1.8. UNIVERSO Y MUESTRA	23
1.8.1. UNIVERSO.....	23
1.8.2. MUESTRA.....	23
1.9. PROPUESTA	23
1.10. ORGANIZACIÓN DE LA TESINA	23
CAPITULO II. MARCO REFERENCIAL	25
2.1. INTELIGENCIA DE NEGOCIOS	26
2.1.1. DEFINICIÓN.....	26

2.1.2.	<i>¿QUIÉNES NECESITAN DE LA INTELIGENCIA DE NEGOCIOS?</i>	27
2.1.3.	<i>BENEFICIOS QUE APORTA LA INTELIGENCIA DE NEGOCIOS</i>	29
2.1.4.	<i>COMPONENTES DE INTELIGENCIA DE NEGOCIOS</i>	29
2.1.5.	<i>ARQUITECTURA DE LA INTELIGENCIA DE NEGOCIOS</i>	30
2.2.	DATAWAREHOUSE	32
2.2.1.	<i>DEFINICION</i>	32
2.2.2.	<i>CARACTERÍSTICAS DEL DATAWAREHOUSE</i>	33
2.2.3.	<i>PROCESOS PARA LA CONSTRUCCIÓN DE UN DATAWAREHOUSE</i>	36
2.3.	DATAMART	37
2.3.1.	<i>DEFINICIÓN DE UN DATAMART</i>	37
2.3.2.	<i>CLASIFICACIÓN DEL DATAMART</i>	38
2.3.3.	<i>TIPOS DE DATAMARTS</i>	39
2.3.4.	<i>DIFERENCIAS ENTRE UN DATAWAREHOUSE Y DATAMART</i>	40
2.4.	BUSINESS INTELLIGENCE ROADMAP	41
2.4.1.	<i>DEFINICION</i>	41
2.4.2.	<i>PASOS DEL BUSINESS INTELLIGENCE ROADMAP</i>	41
CAPITULO III. ESTADO DEL ARTE METODOLÓGICO		43
3.1.	MODELO DE DATAWAREHOUSE PARA EL SECTOR PÚBLICO (EDITORIA PERÚ S.A.)	
	44	
3.1.1.	<i>RESUMEN</i>	44
3.1.2.	<i>DEFINICION DEL PROBLEMA</i>	45
3.1.3.	<i>LIMITACIONES DE LA SOLUCIÓN DEL PROBLEMA</i>	46
3.1.4.	<i>VARIANTES DE LA SOLUCIÓN DEL PROBLEMA</i>	46
3.1.5.	<i>JUSTIFICACIÓN E IMPORTANCIA</i>	47
3.1.6.	<i>OBJETIVOS GENERALES</i>	47
3.1.7.	<i>OBJETIVOS ESPECIFICOS</i>	47
3.1.8.	<i>MODELO OLAP PARA UNA EMPRESA PÚBLICA</i>	47
3.1.9.	<i>EXPERIMENTOS Y RESULTADOS</i>	49
3.2.	MODELO DE DATAWAREHOUSE PARA LA CAJA MURCIA	53
3.2.1.	<i>RESUMEN</i>	53
3.2.2.	<i>DEFINICIÓN DEL PROBLEMA</i>	54
3.2.3.	<i>OBJETIVOS GENERALES</i>	54
3.2.4.	<i>PLANTEAMIENTO DE LA SOLUCION</i>	54
3.2.5.	<i>RESULTADOS</i>	55
3.2.6.	<i>CONCLUSIONES</i>	56
3.3.	MODELO DE DATAWAREHOUSE PARA EL GRUPO SANTO DOMINGO	56
3.3.1.	<i>RESUMEN</i>	56
3.3.2.	<i>DEFINICION DEL PROBLEMA</i>	57
3.3.3.	<i>OBJETIVOS</i>	57
3.3.4.	<i>PROPUESTA</i>	57
3.4.	POSIBLES HERRAMIENTAS A UTILIZAR	57
3.4.1.	<i>SSIS</i>	57
3.4.2.	<i>DATA STAGE</i>	58
3.4.3.	<i>SUNOPSIS</i>	58
3.4.4.	<i>MICROSTRATEGY</i>	58

3.4.6.	<i>BUSSINESS OBJECTS</i>	59
3.4.7.	<i>PENTAH0</i>	59
3.4.8.	<i>OCTOPUS</i>	59
3.5.	CUADRO COMPARATIVO DE LAS HERRAMIENTAS	59
3.5.1.	<i>DEFINICION DE LAS VARIABLES DE COMPARACION</i>	59
3.5.2.	<i>CRITERIO DE PUNTUACION</i>	60
3.5.3.	<i>PUNTUACION DE LAS HERRAMIENTAS</i>	60
3.6.	POSIBLES METODOLOGIAS A UTILIZAR	62
3.6.1.	<i>METODOLOGIA ROADMAP</i>	62
3.6.2.	<i>METODOLOGÍA DE KIMBALL</i>	63
3.6.3.	<i>METODOLOGIA DATAWAREHOUSE ENGINEERING PROCESS (DWEP)</i>	63
3.7.	CUADRO COMPARATIVO DE METODOLOGIAS	64
3.7.1.	<i>DEFINICION DE LAS VARIABLES DE COMPARACION</i>	64
3.7.2.	<i>CRITERIO DE PUNTUACION</i>	64
3.7.3.	<i>PUNTUACION DE LAS METODOLOGIAS</i>	65
CAPITULO IV. IMPLEMENTACIÓN Y RESULTADOS		67
4.1.	DEFINICION DE LAS HERRAMIENTAS A UTILIZAR	68
4.1.1.	<i>MICROSOFT SQL SERVER 2005 ANALYSIS SERVICES (SSAS)</i>	68
4.1.2.	<i>MICROSOFT SQL SERVER 2005 INTEGRATION SERVICES (SSIS)</i>	68
4.1.3.	<i>MICROSOFT SQL SERVER 2005 REPORTING SERVICES (SSRS)</i>	68
4.2.	DEFINICION DE LA METODOLOGIA ELEGIDA	68
4.2.1.	<i>PASOS PARA LA IMPLEMENTACION DEL DATAMART COMO SOLUCION DE INTELIGENCIA DE NEGOCIOS SEGÚN BUSINESS INTELLIGENCE ROADMAP</i>	69
4.3.	APLICACIÓN DE LA GUIA BUSINESS INTELLIGENCE ROADMAP	71
4.3.1.	<i>EVALUACIÓN DEL NEGOCIO</i>	71
4.3.2.	<i>DEFINICIÓN DE REQUERIMIENTOS</i>	73
4.3.3.	<i>ANÁLISIS DE DATOS</i>	74
4.3.4.	<i>PROTOTIPO DE APLICACIÓN</i>	75
4.3.5.	<i>DISEÑO DE LA BASE DE DATOS</i>	76
4.3.6.	<i>DISEÑO DEL ETL</i>	84
4.3.7.	<i>DESARROLLO DEL ETL</i>	86
CAPITULO V. CONCLUSIONES Y RECOMENDACIONES		88
5.1	CONCLUSIONES	89
5.2	RECOMENDACIONES	89
REFERENCIAS BIBLIOGRÁFICAS		90

LISTA DE FIGURAS

Figura 1 Arquitectura de Inteligencia de Negocios [3]	31
Figura 2 Arquitectura de Data Warehouse [4]	33
Figura 3 Data Warehouse y Datos de la Organización [2]	37
Figura 4 Clasificación del DataMart [*]	38
Figura 5 Pasos de Metodología RoadMap	42
Figura 6 Situación actual de Editora Perú S.A.....	46
Figura 7 Modelo OLAP general.....	48
Figura 8 Modelo OLAP General detallado.....	49
Figura 9 Ventas Anuales.....	51
Figura 10 Comparativo Mensual de ventas	51
Figura 11 Solución al problema	52
Figura 12 Cuadro comparativo de herramientas de inteligencia de negocios	62
Figura 13 Cuadro comparativo de metodologías de inteligencia de negocios	66
Figura 14 Metodología adaptada de Business Intelligence RoadMap.....	69
Figura 15 Modelo Físico del proceso de Cotización	77
Figura 16 Modelo Físico del proceso de Entrada y Salida	78
Figura 17 Creación del Origen de Datos para el DataMart.....	79
Figura 18 Selección de tablas para la creación del Origen de Datos	79
Figura 19 Vista del Origen de Datos.....	80
Figura 20 Creación del Cubo con el Origen de Datos.....	80
Figura 21 Creación del Cubo con el Origen de Datos.....	81
Figura 22 Creación de la tabla de hechos con sus dimensiones	81
Figura 23 Selección de tabla de hechos y sus dimensiones	82
Figura 24 Creación de Dimensiones del Cubo.....	82
Figura 25 Creación del Cubo con sus Dimensiones.....	83
Figura 26 Creación del Cubo	83
Figura 27 Proceso de Extracción, transformación y Carga	85
Figura 28 Plataforma de Inteligencia de negocios.....	85
Figura 29 Plataforma de Aplicación de Soluciones.....	85
Figura 30 Ejemplo de Consultas.....	86
Figura 31 Generación de Reportes	87

LISTA DE TABLAS

Tabla 0.1 Diferencia entre Sistemas transaccionales y sistemas basados en DataMart.....	15
Tabla 3.1 Definición y ponderación de variables de Comparación de herramientas BI	60
Tabla 3.2 Cuadro de puntuación de las herramientas BI.....	60
Tabla 3.3 Cuadro comparativo de herramientas BI	61
Tabla 3.4 Definición y ponderación de variables de Comparación de las metodologías	64
Tabla 3.5 Cuadro de puntuación de las metodologías BI	65
Tabla 3.6 Cuadro comparativo de metodologías BI.....	65

INTRODUCCION

Desde el inicio de la Era de la Información las empresas necesitan explotar su mayor recurso, la información. La explotación eficiente de la información permite una rápida, acertada y oportuna toma de decisiones bajo el manejo de datos confiables.

Con el crecimiento de las empresas se ha elevado a gran medida la capacidad de generación y almacenamiento de la información, que no puede ser analizada por los métodos tradicionales existentes. Mientras mayor es la capacidad para almacenar datos, mayor es la incapacidad para extraer información realmente útil. Mucha información importante para la toma de decisiones queda oculta y muchas veces no tomadas en cuenta debido a que no se pudo obtener esa valiosa información en el momento requerido, así como también los sistemas actuales no están preparados para enfrentar la gran magnitud de la información.

Por esta razón muchas organizaciones han logrado implementar un DataMart que permita centralizar la información útil y necesaria para ayudar a la toma de decisiones en los siguientes niveles organizacionales:

- ✚ Nivel operativo,

- ✚ Nivel táctico y

- ✚ Nivel estratégico.

El presente trabajo se orienta a implementar un DataMart como una solución para aplicar Inteligencia de Negocios en el área de Logística de T-IMPULSO, empresa encargada de brindar servicios de Recursos Humanos principalmente.

El DataMart está basado en un procesamiento distinto al utilizado por los sistemas operacionales, es decir se basa en Procesos de Análisis en Línea (OLAP – Online

Analysis Process), usado en el análisis de negocios y otras aplicaciones que requieren una visión flexible del negocio.

A continuación presentamos algunas principales diferencias entre los sistemas transaccionales y los sistemas basados en DataMart.

Sistemas Transaccionales	Sistemas basados en DataMart
<ul style="list-style-type: none"> Admiten el acceso simultáneo de muchos usuarios que agregan y modifican datos. Predomina el proceso puntual, es decir, trabaja con base en transacciones Orientado hacia las transacciones Orientado hacia las operaciones 	<ul style="list-style-type: none"> Admiten el acceso simultáneo de muchos usuarios que consultan mas no modifican datos. Predomina el proceso masivo. Se trabaja con base en resúmenes, Orientado al análisis Orientado hacia conceptos de negocio
<ul style="list-style-type: none"> Representa el estado, en cambio constante, de una organización, pero no guardan su historia. 	<ul style="list-style-type: none"> Guardan el historial de una organización.
<ul style="list-style-type: none"> Contienen grandes cantidades de datos, incluyendo los datos intensivos. 	<ul style="list-style-type: none"> Contienen grandes cantidades de datos, resumizados, consolidados y transformados

<ul style="list-style-type: none"> ✚ Datos en general desagregados, detallados 	<ul style="list-style-type: none"> ✚ Datos en distintos niveles de detalle y agregación, totalizados.
<ul style="list-style-type: none"> ✚ Tienen estructura de base de datos compleja. ✚ Estructura relacional ✚ Estructura estática 	<ul style="list-style-type: none"> ✚ Tienen estructura de base de datos ✚ Visión multidimensional ✚ Visión flexible

Tabla 0.1 Diferencia entre Sistemas transaccionales y sistemas basados en DataMart.

Fuente: Elaboración propia

CAPITULO I. PLANTEAMIENTO METODOLÓGICO

1.1. ANTECEDENTES

Desde el inicio de la Era de la información mayor ha sido la necesidad de procesar datos a través de Sistemas de Información u otros medios automatizados, resulta complicado tomar decisiones de manera rápida, oportuna y confiable si no se tienen todos los datos centralizados y homogenizados.

Debido a que los volúmenes de datos que se deben procesar ya alcanzan tamaños considerables, y que en la mayoría de los casos no necesariamente éstos son manipulados por un solo sistema de información u otro tipo de herramienta resulta muy común que se tenga que trabajar con ambientes no homogéneos. Es por ello que muchas de las organizaciones desde hace ya algunos años han venido estudiando nuevas alternativas para el procesamiento de datos altamente voluminoso.

La empresa T-IMPULSO atraviesa por un proceso de crecimiento por la cual maneja una gran cantidad de información día a día. T-Impulso es una empresa que se dedica a la tercerización y, outsourcing de personal como también brinda servicios de nóminas, actualmente cuenta con unos 6000 trabajadores entre Empleado y Obreros; entre los principales Clientes están Gloria, Mineras San Martín (que contiene un grupo importante de mineras del entorno nacional), de la misma manera contamos con cierta cantidad de proveedores que nos brindan las herramientas, servicios y productos necesarios que hacen que nuestro servicio sea más efectivo. Actualmente los usuarios del área de Logística de la empresa T-Impulso, realizan sus actividades de forma ineficiente, esto se debe en gran medida al volumen de información que crece día a día y a que los datos se encuentran dispersos en varias fuentes y formatos (ya sea en archivos digitales, en documentos y en la base de datos transaccional diaria), lo que ocasiona que su recolección, homogenización y análisis se conviertan en tareas tediosas y que a su vez conlleva a desperdiciar tiempo en actividades que no generan valor (esto se da en el caso de la recolección y homogenización de datos)

1.2. DEFINICION DEL PROBLEMA

Los tomadores de Decisiones de T Impulso siendo uno de ellos el jefe del Área Logística requieren reportes resumidos para la toma de decisiones a mediano y corto plazo. Para cumplir tales requerimientos actualmente los empleados del área de logística realizan reportes manualmente desde los archivos que tienen en su poder (Excel, documentos digitalizados y documentos físicos), también se ayudan de las diferentes tecnologías que les ofrece la empresa (Sistema Transaccional) y esto ocasiona que se tengan resultados en diferentes formatos y no tan certeros, ya que sabemos que la información transaccional puede cambiar de un momento a otro. Por ende el problema principal radica en que el proceso de extracción e integración de los datos para la realización de los reportes se realiza manualmente y sistemáticamente, resultando muy complicado ,provocando retraso en la atención de requerimientos haciendo que en muchos casos la entrega de la información no sea oportuna; por otro lado la elaboración de algunos reportes requiere mayor conocimiento en el manejo de las herramientas informáticas, por lo que son derivados al área de Tecnología de Información, ocasionando mas pérdida de tiempo

El problema de falta de homogenización de las fuentes no lo solucionaremos ya que resulta común que se trabaje con archivos y base de datos en diferentes tecnologías y formatos. Nuestra solución se limita a extraer datos de las fuentes del área de logística, sintetizar, centralizar e integrarla para resolver los problemas antes mencionados como la entrega de reportes con información precisa y a corto plazo.

De esta manera se explotará eficientemente la información del área de logística.

1.3. FORMULACION DEL PROBLEMA

1.3.1. PROBLEMA GENERAL

La demora en la generación de reportes solicitados por los tomadores de decisiones del área de Logística.

1.3.2. PROBLEMA ESPECIFICO 1

Falta de integración de datos que son relevantes para la toma de decisiones del área de Logística de T-IMPULSO.

1.4. OBJETIVOS

1.4.1. OBJETIVO GENERAL

La implementación de un DataMart como solución para reducir los tiempos en la elaboración de los reportes solicitados por los tomadores de decisiones del área de Logística.

1.4.2. OBJETIVO ESPECIFICO 1

Integrar y automatizar la recolección de la información del área de logística

1.5. JUSTIFICACION

1.5.1. POBLACIÓN A LA QUE BENEFICIA LA INVESTIGACIÓN

La población a la que beneficiará la presente investigación son:

1. La empresa T - IMPULSO, con un beneficio económico debido a que con el desarrollo del presente proyecto obtendrá ventaja competitiva, como resultado de las decisiones oportunas que vaya a tomar el área de Logística con respecto a las compras seleccionando adecuadamente a los proveedores.
2. Los proveedores, debido a que con el desarrollo del presente proyecto, el área de Logística podrá redirigir mejor las compras de la empresa y de esta manera hacerle saber al proveedor el comportamiento de las compras en el tiempo oportuno.

1.5.2. JUSTIFICACIÓN DEL PROYECTO

1.5.2.1. Justificación Tecnológica

La empresa T-Impulso logrará estar a la par de los competidores Nacionales como con los competidores Transnacionales con respecto a las tecnologías que usan.

1.5.2.2. Justificación Económica

La empresa al obtener la información más integrada podrá reducir los tiempos al momento de que se le solicite algún tipo de información y esto se verá reflejada al momento de realizar un reporte, debido a ello consumiríamos menos tiempo y ahorraríamos en los costos.

1.5.3. FACTIBILIDAD DEL ESTUDIO

El desarrollo del presente proyecto es factible, porque va acompañado del apoyo administrativo de la empresa T - Impulso y su área de Logística, brindándonos acceso y autorización para manejar bases de datos del área.

Además de lo mencionado en el párrafo anterior, la factibilidad del proyecto también está influenciada negativamente por los siguientes factores:

El corto tiempo que nos brindan los trabajadores del área de Logística es corto, siendo necesaria su colaboración para realizar el análisis de requerimientos de datos.

Las pocas reuniones que podamos acordar con los jefes de las áreas objetivo.

Menor resistencia al cambio de los usuarios que puede ser bloqueante para el proyecto.

Utilizar una herramienta de SQL Express por ser libre.

El corto tiempo que tenemos para el desarrollo del DataMart, teniendo en cuenta que se debe satisfacer las expectativas de los usuarios.

Acceso limitado a la información

1.5.4. IMPACTO ESPERADO DE LOS RESULTADOS

El principal impacto que tendrá el desarrollo del presente proyecto sería el apoyo en la toma de decisiones del área de Logística de la empresa, de forma que se podrá redirigir mejor los esfuerzos para brindar las compras y servicios adecuados para nosotros y poder darles el uso adecuado, realizando esto en menos tiempo y de forma más precisa, teniendo facilidad en la recolección, homogeneización y presentación de la información para su posterior análisis.

Otro aspecto que se tiene que tener en cuenta dentro del impacto, sería la construcción del DataMart, que permitirá llevar a la organización a la par tecnológica de la competencia, mejorando su posición en el mercado.

1.5.5. OPORTUNIDAD PARA TRATAR EL PROBLEMA

Al desarrollar el presente proyecto además de buscar desarrollar los objetivos mencionados, también se tiene la oportunidad de mejorar:

La forma como trabaja el personal del área de Logística, debido a que no gastarían esfuerzos en la definición, recolección y transformación de la data relevante para la toma de decisiones, sabiendo que esto estaría automatizado por el DataMart.

Exclusión de datos que pueden ser considerados como ruido para el análisis, esto se daría luego de la definición de los datos y fuentes de datos relevantes para el análisis y la toma de decisiones.

1.6. DELIMITACION DE LA INVESTIGACION

1.6.1. DELIMITACIÓN DEL PROBLEMA

En el proyecto a desarrollar se considera al área de Logística de la empresa T - Impulso como foco de estudio, análisis de requerimientos de información y análisis de entrega de resultados.

1.6.2. ASPECTOS DE ESTUDIO QUE COMPRENDE EL PROBLEMA

Para tratar el problema definido en párrafos anteriores, se tiene que abordar los siguientes aspectos:

Logística

Recursos Humanos.

El almacenamiento de grandes volúmenes de datos

La transformación de los datos para un análisis homogeneizado.

Definición de requerimientos de información de áreas empresariales

1.7. TIPO DE INVESTIGACIÓN AL QUE PERTENECE EL ESTUDIO

1.7.1. UNIDAD DE ANÁLISIS

La unidad a analizar será el área de Logística de la Empresa T - Impulso.

1.7.2. JUSTIFICACIÓN DEL TIPO DE INVESTIGACIÓN QUE SE PROPONE

El tipo de investigación a la que pertenece el presente proyecto de Tesina es el de Investigación Aplicada, puesto que se pretende la construcción de una aplicación (DataMart) que realizará varios procesos sobre los datos (recolección, homogeneización, integración y presentación), de manera que pueda generar información relevante para la toma de decisiones del área de Logística, y así poder darle solución al problema en cuestión.

1.8. UNIVERSO Y MUESTRA

1.8.1. UNIVERSO

El universo está conformado por todas las fuentes de datos que utiliza el área de logística en las Sucursales de Lima, Arequipa y Trujillo, para realizar actividades que brinden información sobre los proveedores y las compras de la empresa.

1.8.2. MUESTRA

La muestra está conformada por todas las compras que realiza la empresa en Lima por sus usuarios del área de logística.

1.9. PROPUESTA

La solución propuesta en este trabajo es la implementación de un DataMart para el área de logística de T-IMPULSO, que implica la transformación, homogenización e integración de datos para brindarles una herramienta que les facilitará la toma de decisiones. Los usuarios internos (jefes, analistas, supervisores) podrán desarrollar sus análisis, consultas, seguimientos, etc. De manera independiente y sin necesidad de conocimientos técnicos.

1.10. ORGANIZACIÓN DE LA TESINA

En el Capítulo II. Marco Referencial se describen los conceptos y tecnologías que establecen el contexto de trabajo de la presente tesis, principalmente aquellos que están relacionados a DataMart, Data Warehouse e Inteligencia de Negocios, de forma que se brinda las bases teóricas para el sustento del desarrollo de la tesina. En el Capítulo III. Estado del Arte Metodológico se muestra casos de éxito en la implementación de un DataMart en algunas empresas como la descripción de las metodologías y herramientas a utilizar. En el Capítulo IV. Implementación y Resultados se estructura el diseño de la implementación listando las herramientas y actividades que se deben utilizar y realizar para la implementación del DataMart. El Capítulo V. Discusión, Conclusiones y Recomendaciones comprenderá las

conclusiones y del desarrollo de la tesina y algunas recomendaciones para posteriores trabajos basados en esta tesina.

CAPITULO II. MARCO REFERENCIAL

2.1. INTELIGENCIA DE NEGOCIOS

2.1.1. DEFINICIÓN

El concepto de Inteligencia de Negocios no es un resultado de desarrollos en el mundo de las Ciencias Administrativas, sino que es un producto del progreso de la informática o de la denominada “infotecnología”. [1]

Tampoco es un concepto nuevo originado ya que su origen data de la publicación en el IBM Journal de 1958, del artículo de *Hans Peter Luhn* titulado “A Business Intelligence System” donde se define con detalle el concepto con una perspectiva, que solo en nuestros días, ha sido posible su plena utilización. [1]

La aplicación amplia de este concepto, que ha generado el desarrollo de un mercado importante de productos de software alrededor de él, se ha hecho posible gracias a los avances de la tecnología y a que los ejecutivos de las empresas han entendido que el acceso rápido y oportuno al conocimiento empírico sobre su negocio, representa mejoras substanciales en los resultados [1].

Inteligencia de negocios es un proceso interactivo para explorar y analizar información estructurada sobre un área (normalmente almacenada en un Data Warehouse), para descubrir tendencias o patrones, a partir de los cuales derivar ideas y extraer conclusiones. [1]

Es la combinación de tecnología, herramientas y procesos que permiten transformar datos almacenados en información, esta información en conocimiento y este conocimiento dirigido a un plan o una estrategia comercial. La inteligencia de negocios debe ser parte de la estrategia empresarial, esta le permite optimizar la utilización de recursos, monitorear el cumplimiento de los objetivos de la empresa y la capacidad de tomar buenas decisiones para así obtener mejores resultados. [29]

El proceso de Inteligencia de Negocios incluye la comunicación de los descubrimientos y efectuar los cambios.

Las áreas incluyen clientes, proveedores, productos, servicios y competidores.
[2]

Detallaremos los siguientes términos:

Proceso interactivo: Análisis de información continuado en el tiempo, de esta manera podemos ver tendencias, cambios, variabilidades, etc.[3]

Explorar: Acceder a información que nos facilite su interpretación, de esta manera comprender que sucede en nuestro negocio; es posible que se encuentren nuevas relaciones que hasta el momento se desconocía. [3]

Analizar: Descubrir relaciones entre variables, como tendencias, es decir, cual puede ser la evolución de la variable, o patrones. Por ejemplo si un Cliente tiene ciertas características, cual es la probabilidad que otro cliente con características similares actúe igual que el anterior cliente. [3]

Información estructurada y Data Warehouse: La información que se encuentra almacenada en las tablas relacionadas del Data Warehouse. [3]

El origen de la Inteligencia de Negocios va ligado a proveer acceso directo a la información a los usuarios de negocio para ayudarles en la toma de decisiones, sin intervención de los Sistemas de Información. [3]

2.1.2. ¿QUIÉNES NECESITAN DE LA INTELIGENCIA DE NEGOCIOS?

Todas aquellas personas de las organizaciones que tengan que tomar decisiones. Dependiendo de qué preguntas necesiten responder, se establece el Modelo de Inteligencia de Negocios necesario. [3]

De esta forma las decisiones a tomar pueden ser de tipo operativo, táctico o estratégico. [3]

2.1.2.1. Inteligencia de negocios a Nivel Estratégico

Permite que la alta dirección de las empresas pueda analizar y monitorear tendencias, patrones, metas y objetivos estratégicos de la organización. Un ejemplo de Inteligencia de Negocios a nivel estratégico lo constituye el Cuadro de Mando Integral o BalancedScorecard, concepto introducido por Robert Kaplan y David Norton el cual definen como:

"Un esquema de trabajo multidimensional para describir, implementar y administrar estrategia a todo nivel dentro de una empresa, a través de la vinculación de objetivos, iniciativas y mediciones a la estrategia de la organización [4].

2.1.2.2. Inteligencia de negocios a Nivel Táctico

Permite que los analistas de datos y la gerencia media de la empresa utilicen herramientas de análisis y consulta con el propósito de tener acceso a la información sin intervención de terceros. [5].

Como ejemplo un gerente de ventas recibe un reporte pre impreso en donde se indica que las ventas de una determinada categoría de productos o servicios, se incrementaron de manera inusual con relación al periodo anterior, una herramienta de análisis y consulta le permite analizar éste incremento y establecer si el mismo se debe a nuevos productos, nuevos proveedores o una estrategia de promociones que haya producido el incremento en la demanda. [5].

2.1.2.3. Inteligencia de negocios a Nivel operativo

Permite que los empleados que trabajan con información operativa puedan recibir la misma información de una manera oportuna, exacta y adecuada y se componen básicamente de herramientas de reportes u hojas de cálculo con un formato fijo cuya información se actualiza frecuentemente [5].

Un ejemplo de esto podría ser un supervisor de ventas que utiliza una hoja de cálculo para monitorear el cumplimiento de las cuotas de ventas de los

vendedores a su cargo, una de las columnas tendría una información fija (la cuota de ventas) y a su lado podría estar una columna que diariamente extraiga el total de ventas para ése vendedor en particular. El supervisor de ventas a su vez podría aplicar fórmulas tomando en cuenta la columna de cuota y la columna de venta real sin necesidad de tener que introducirlas de manera manual [5].

2.1.3. BENEFICIOS QUE APORTA LA INTELIGENCIA DE NEGOCIOS

Los beneficios que brinda la Inteligencia de Negocios pueden ser de distintos tipos:

Beneficios tangibles: Reducción de costes, generación de ingresos, reducción de tiempos para las distintas actividades del negocio. [6].

Beneficios intangibles: Obtener la información disponible para la toma de decisiones hará que más usuarios utilicen dicha información para mejorar nuestra posición competitiva. [6].

Beneficios estratégicos: Todos aquellos que nos facilitan la formulación de la estrategia, es decir, a que proveedores, mercados y con qué productos dirigirnos. [6].

2.1.4. COMPONENTES DE INTELIGENCIA DE NEGOCIOS

Todas las soluciones de Inteligencia de negocios tienen funciones parecidas, pero deben de reunir al menos los siguientes componentes:

Multidimensionalidad: La información multidimensional se puede encontrar en hojas de cálculo, bases de datos, etc. Una herramienta de Inteligencia de negocios debe de ser capaz de reunir información dispersa en toda la empresa e incluso en diferentes fuentes para así proporcionar a los departamentos la accesibilidad, poder y flexibilidad que necesitan para analizar la información. Por ejemplo, un pronóstico de ventas de un nuevo producto en varias regiones no está completo si no se toma

en cuenta también el comportamiento histórico de las ventas de cada región y la forma en que la introducción de nuevos productos se ha desarrollado en cada región en cuestión.

Data Mining: Las empresas suelen generar grandes cantidades de información sobre sus procesos productivos, desempeño operacional, mercados y proveedores. Pero el éxito de los negocios depende por lo general de la habilidad para ver nuevas tendencias o cambios en las tendencias. Las aplicaciones de data mining pueden identificar tendencias y comportamientos, no sólo para extraer información, sino también para descubrir las relaciones en bases de datos que pueden identificar comportamientos que no fueron muy evidentes.

Agentes: Los agentes son programas que "piensan". Ellos pueden realizar tareas a un nivel muy básico sin necesidad de intervención humana. Por ejemplo, un agente pueden realizar tareas un poco complejas, como elaborar documentos, establecer diagramas de flujo, etc.

Data Warehouse: Es la respuesta de la tecnología de información a la descentralización en la toma de decisiones. Coloca información de todas las áreas funcionales de la organización en manos de quien toma las decisiones. También proporciona herramientas para búsqueda y análisis.[6]

2.1.5. ARQUITECTURA DE LA INTELIGENCIA DE NEGOCIOS

En la figura 1 se muestra la arquitectura de la inteligencia de negocios.

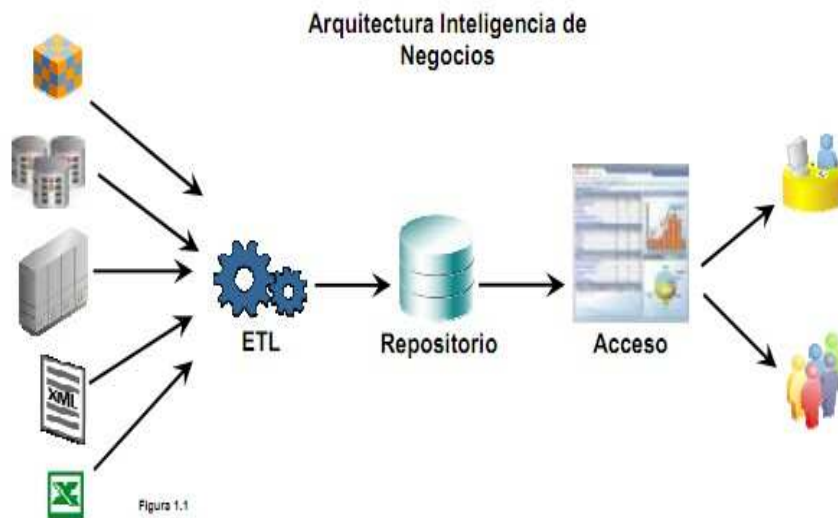


Figura 1 Arquitectura de Inteligencia de Negocios [3]

2.1.5.1. Elementos de la arquitectura de la Inteligencia de negocio

Fuentes de datos: Representan las distintas fuentes de datos (Cubos, bases de datos, Sql Server, mainframe, archivos planos, archivos xml, hojas de Excel, etc.) que pudieran utilizarse para extraer los datos de múltiples fuentes simultáneamente.

ETL: Proceso de extracción transformación y carga. Este proceso es en el que se definen de las fuentes heterogéneas que campos se van a utilizar, si necesitan algún tipo de modificación y/o transformación y donde quiero ubicar estos datos, este proceso se le conoce como “mapping”.

Repositorio: Aquí es donde se encuentran los datos transformados representados visualmente en modelos multidimensionales, dimensiones y tablas de datos.

Acceso: Existe un proceso entre el repositorio de datos y la interfase de acceso al usuario, este es el motor de Inteligencia de Negocios que me permite habilitar componentes, administrar consultas, monitorea procesos, cálculos y métricas.

La interface de acceso a usuarios permite interaccionar con los datos, representar de forma gráfica con aquellos resultados de las consultas y los indicadores de gestión que fueron construidos.[3]

2.2. DATAWAREHOUSE

2.2.1.DEFINICION

Un Data Warehouse es una base de datos corporativa que se caracteriza por integrar y depurar información de una o más fuentes distintas, para luego procesarla permitiendo su análisis desde infinidad de perspectivas y con grandes velocidades de respuesta. Por lo general, mantiene años de historia y se consulta para la inteligencia de negocios o de otras actividades analíticas. La creación de un Data Warehouse representa en la mayoría de las ocasiones el primer paso, desde el punto de vista técnico, para implantar una solución completa y fiable de Business Intelligence.

La ventaja principal de este tipo de bases de datos radica en las estructuras en las que se almacena la información (modelos de tablas en estrella, en copo de nieve, cubos relacionales... etc.) que más adelante lo comentaremos. Este tipo de persistencia de la información es homogénea y fiable, y permite la consulta y el tratamiento jerarquizado de la misma (siempre en un entorno diferente a los sistemas operacionales). [3]

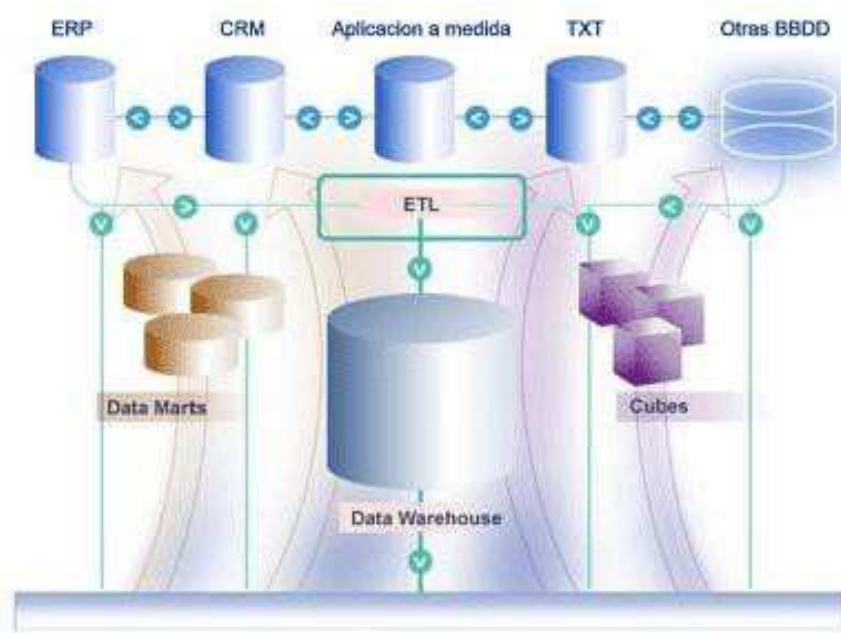


Figura 2 Arquitectura de Data Warehouse [4]

2.2.2. CARACTERÍSTICAS DEL DATAWAREHOUSE

El término Data Warehouse fue acuñado por primera vez por Bill Inmon, y se traduce literalmente como *almacén de datos*. No obstante, y como cabe suponer, es mucho más que eso. Según definió el propio Bill Inmon, un Data Warehouse se caracteriza por ser:

Integrado: Los datos almacenados en el Data Warehouse deben integrarse en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas operacionales deben ser eliminadas. La información suele estructurarse también en distintos niveles de detalle para adecuarse a las distintas necesidades de los usuarios.

Temático: Sólo los datos necesarios para el proceso de generación del conocimiento del negocio se integran desde el entorno operacional. Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales. Por ejemplo, todos los datos sobre proveedores pueden ser consolidados en una única tabla del Data Warehouse. De esta forma, las peticiones de información sobre

proveedores serán más fáciles de responder dado que toda la información reside en el mismo lugar.

Histórico: El tiempo es parte implícita de la información contenida en un Data Warehouse. En los sistemas operacionales, los datos siempre reflejan el estado de la actividad del negocio en el momento presente. Por el contrario, la información almacenada en el Data Warehouse sirve, entre otras cosas, para realizar análisis de tendencias. Por lo tanto, el Data Warehouse se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones.

No volátil: El almacén de información de un Data Warehouse existe para ser leído, pero no modificado. La información es por tanto permanente, significando la actualización del Data Warehouse la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre lo que ya existía.[3]

Otra característica del Data Warehouse es que contiene metadatos, es decir, datos sobre los datos. Los metadatos permiten saber la procedencia de la información, su periodicidad de refresco, su fiabilidad, forma de cálculo, etc. Los metadatos serán los que permiten simplificar y automatizar la obtención de la información desde los sistemas operacionales a los sistemas informacionales.

Objetivos de los Metadatos:

Dar soporte al usuario final, ayudándole a acceder al Data Warehouse con su propio lenguaje de negocio, indicando qué información hay y qué significado tiene. Ayudar a construir consultas, informes y análisis, mediante herramientas de Inteligencia de Negocios como DSS, EIS o CMI.

DSS - Sistema de Soporte de Decisiones

Según [IBM 1999], un Sistema de Soporte de Decisiones (*DSS-Decision Support System*) contiene todos los servicios y procesos, para seleccionar, manipular, y analizar información y presentar resultados. Debe de permitir acceso transparente a la data en varias partes del Data Warehouse y proveer una interfaz común para los diferentes grupos de usuarios. Un DSS también puede ser definido como un sistema computacional diseñado para apoyar en los procesos de la toma de decisiones en una organización. Un DSS es la ventana del usuario a los datos almacenados en el ambiente del Data Warehouse.

Sistemas de información ejecutiva (EIS).

Los sistemas de información ejecutiva (EIS, *Executive Information System*) son el tipo de DSS que más se suele emplear en Business Intelligence, ya que proveen a los gerentes de un acceso sencillo a información interna y externa de su compañía, y que es relevante para sus factores clave de éxito.

Cuadro de Mando Integral – CMI

Fue presentado en el número de enero/febrero de 1992 de la revista Harvard Business Review, con base en un trabajo realizado para una empresa de semiconductores. Sus autores, Robert Kaplan y David Norton, plantean que el CMI es un sistema de administración o sistema administrativo (Management System), que va más allá de la perspectiva financiera con la que los gerentes acostumbran evaluar la marcha de una empresa.

Es un método para medir las actividades de una compañía en términos de su visión y estrategia. Proporciona a los gerentes una mirada global del desempeño del negocio.

Es una herramienta de administración de empresas que muestra continuamente cuándo una compañía y sus empleados alcanzan los

resultados definidos por el plan estratégico. También es una herramienta que ayuda a la compañía a expresar los objetivos e iniciativas necesarias para cumplir con la estrategia.

Dar soporte a los responsables técnicos del Data Warehouse en aspectos de auditoría, gestión de la información histórica, administración del Data Warehouse, elaboración de programas de extracción de la información, especificación de las interfaces para la realimentación a los sistemas operacionales de los resultados obtenidos, etc.[3]

2.2.3. PROCESOS PARA LA CONSTRUCCIÓN DE UN DATAWAREHOUSE

Es importante entender cuál es el proceso de construcción de una Data Warehouse, denominada ETL (Extracción, transformación y carga) a partir de los sistemas operacionales de la organización.

✚ **Extracción:** Obtención de información de las distintas fuentes tanto internas como externas.

✚ **Transformación:** Filtrado, limpieza, depuración, homogeneización y agrupación de la información.

✚ **Carga:** Organización y actualización de los datos y los metadatos en la base de datos.

Una de las claves del éxito en la construcción de un Data Warehouse es el desarrollo de forma gradual, seleccionando a un departamento usuario como piloto y expandiendo progresivamente el almacén de datos a los demás usuarios. Por ello es importante elegir este usuario inicial o piloto, siendo importante que sea un departamento con pocos usuarios, en el que la necesidad de este tipo de sistemas es muy alta y se pueda obtener y medir resultados a corto plazo.[6]

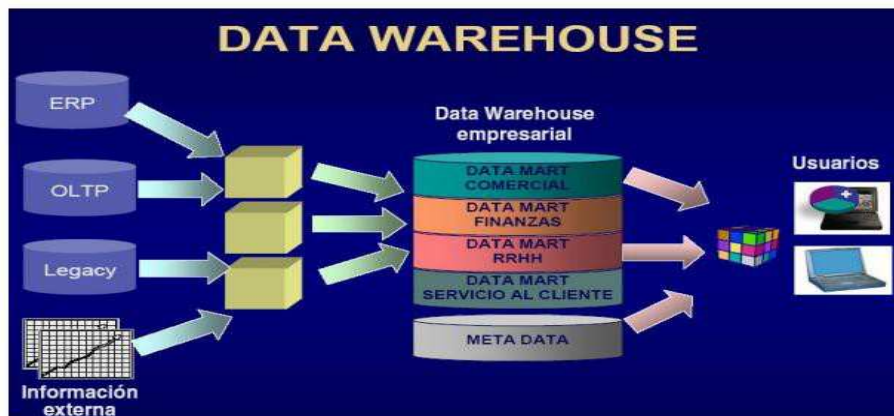


Figura 3 Data Warehouse y Datos de la Organización [2]

2.2.4. PRINCIPALES APORTACIONES DE UN DATAWAREHOUSE

Proporciona una herramienta para la toma de decisiones en cualquier área funcional, basándose en información integrada y global del negocio.

Facilita la aplicación de técnicas estadísticas de análisis y modelización para encontrar relaciones ocultas entre los datos del almacén; obteniendo un valor añadido para el negocio de dicha información.

Proporciona la capacidad de aprender de los datos del pasado y de predecir situaciones futuras en diversos escenarios.

Simplifica dentro de la empresa la implantación de sistemas de gestión integral de la relación con el cliente.

Supone una optimización tecnológica y económica en entornos de Centro de Información, estadística o de generación de informes con retornos de la inversión espectaculares

2.3. DATAMART

2.3.1. DEFINICIÓN DE UN DATAMART

Un **DataMart** es una base de datos departamental, especializada en el almacenamiento de los datos de un área de negocio específica. Se caracteriza por disponer la **estructura óptima de datos** para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento. Un DataMart puede ser alimentado desde los datos de un Data

Warehouse, o integrar por sí mismo un compendio de distintas fuentes de información.

Por tanto, para crear el DataMart de un área funcional de la empresa es preciso encontrar la estructura óptima para el análisis de su información, estructura que puede estar montada sobre una base de datos OLTP, como el propio Data Warehouse, o sobre una base de datos OLAP. La designación de una u otra dependerá de los datos, los requisitos y las características específicas de cada departamento. [7]

2.3.2. CLASIFICACIÓN DEL DATAMART

En la Figura4. Se observa la clasificación del DataMart según Inmon.

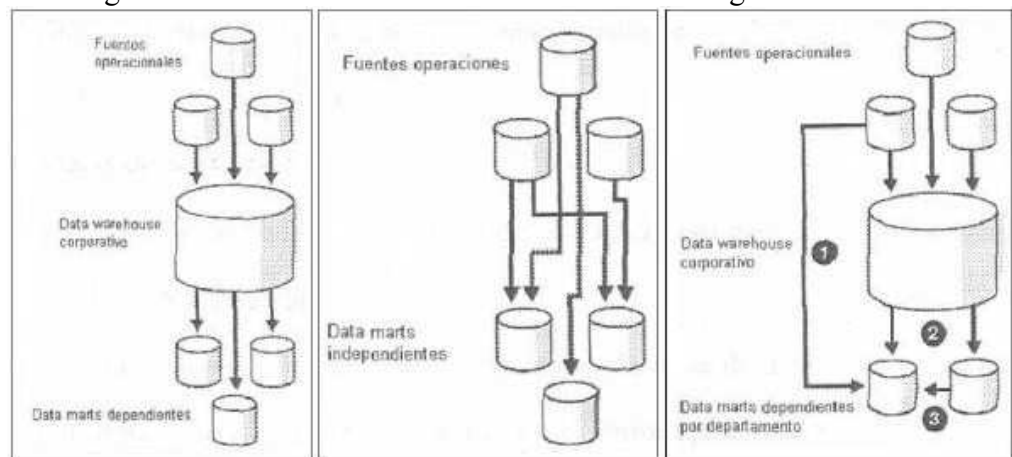


Figura II.1 Datamart dependiente.
Tomado de "Oracle8i Data
Warehousing Guide".

Figura II.2 Datamart independiente.
Tomado de "Oracle8i Data
Warehousing Guide".

Figura II.3 Datamart híbrido.
Tomado de "Oracle8i Data
Warehousing Guide".

Figura 4 Clasificación del DataMart [*]

2.3.2.1. DataMart dependiente

Los DataMart dependientes son aquellos que reciben los datos desde una Data Warehouse. En este tipo de DataMart la fuente de los datos es única, tal como se muestra en la figura 4.1.

2.3.2.2. DataMart independiente

Los DataMart denominados independientes son aquellos que toman sus datos directamente desde los sistemas transaccionales y no dependen de otros

Data Warehouse. Este tipo de DataMart se alimenta generalmente de los denominados “legacy systems” de las organizaciones. Un esquema de este tipo de DataMart se muestra en la figura 4.2

La ventaja de los DataMart dependientes según Inmon (1999) es que son de estructura y arquitectura estable a medida que crece la organización porque se alimentan con el Data Warehouse, el cual crece junto a la organización. Sin embargo, los DataMart independientes manifiestan sus deficiencias posterior a su elaboración cuando existen ya diversos DataMart inconexos.

2.3.2.3. DataMart híbrido

Los DataMart híbridos permiten combinar las fuentes de datos de un Data Warehouse corporativo con otras fuentes de datos tales como sistemas transaccionales y/u operacionales, tal como se muestra en la figura 4.3

2.3.3. TIPOS DE DATAMARTS

DataMart OLAP

Se basan en los populares cubos OLAP, que se construyen agregando, según los requisitos de cada área o departamento, las dimensiones y los indicadores necesarios de cada cubo relacional. El modo de creación, explotación y mantenimiento de los cubos OLAP es muy heterogéneo, en función de la herramienta final que se utilice.

DataMart OLTP

Pueden basarse en un simple extracto del Data Warehouse, no obstante, lo común es introducir mejoras en su rendimiento (las agregaciones y los filtrados suelen ser las operaciones más usuales) aprovechando las características particulares de cada área de la empresa. Las estructuras más comunes en este sentido son las tablas report, que vienen a ser *fact-tables* reducidas (que agregan las dimensiones oportunas), y las vistas materializadas, que se construyen con la misma estructura que las anteriores, pero con el objetivo de explotar la reescritura de queries (aunque sólo es posible en algunos SGBD avanzados, como Oracle).[7]

Los DataMarts que están dotados con estas estructuras óptimas de análisis presentan las siguientes ventajas:

- Poco volumen de datos
- Mayor rapidez de consulta
- Consultas SQL y/o MDX sencillas
- Validación directa de la información
- Facilidad para la historización de los datos

2.3.4. DIFERENCIAS ENTRE UN DATAWAREHOUSE Y DATAMART

En la tabla 2 podemos apreciar las diferencias que existen entre Data Warehouse y DataMart

	Datawarehouse	Datamart
Alcance	Construido para satisfacer las necesidades de información de toda la organización	Construido para satisfacer las necesidades de un área de negocios específica.
Objetivo	Diseñado para optimizar la integración y la administración de los datos fuente	Diseñado para optimizar la entrega de información de soporte a decisiones
Características de los datos	Administra grandes cantidades de datos históricos a nivel atómico	Se concentra en administrar resúmenes y/o datos totalizados
Pertenencia	Pertenece a toda la organización	Pertenece al área de negocio al cual está orientado
Administración	Es administrado por la unidad de sistema de la organización	Es administrado por el personal de sistema de la unidad propietaria del Datamart

Tabla 2.1 Diferencias entre Data Warehouse y DataMart

2.4. BUSINESS INTELLIGENCE ROADMAP

2.4.1.DEFINICION

En términos generales de BI, el Business Intelligence RoadMap (BIR) especifica el camino y la dirección que deben seguir las aplicaciones, estructuras, herramientas y personas que intervienen en un proyecto de este tipo. Este BIR es, en primer lugar, una guía del ciclo de vida de un proyecto para desarrollar aplicaciones de soporte a la decisión utilizando datos estructurados. El 60% de los proyectos de BI son abandonados o acaban fallando por una inadecuada planificación, por falta de tareas, por entregas fuera de plazo, por una mala gestión del proyecto, por una ausencia de requisitos de negocio o por una mala calidad en las entregas. Los gestores necesitan saber qué hacer y no hacer en implementaciones de BI, basándose en experiencias fiables. [26]

En el BIR se definen las actividades necesarias en un proyecto BI para mantener la integración de la infraestructura del entorno al que pertenecen. Las infraestructuras técnicas y no técnicas son las competencias esenciales para la alineación organizacional. Además, se deben definir los roles y responsabilidades asignados a cada persona del equipo para cada paso del desarrollo. BIR describe 16 pasos a seguir en un proyecto de BI, los cuales se encuentran repartidos dentro de los seis estados que se han descrito antes como necesarios para crear un proyecto de ingeniería con éxito. A continuación se van a describir cada uno de ellos dentro de su estado correspondiente. [26]

2.4.2.PASOS DEL BUSINESS INTELLIGENCE ROADMAP

En la figura 5 se muestra los pasos por cada fase de la metodología ROADMAP.



Figura 5 Pasos de Metodología RoadMap

CAPITULO III. ESTADO DEL ARTE METODOLÓGICO

3.1. MODELO DE DATAWAREHOUSE PARA EL SECTOR PÚBLICO

(EDITORIA PERÚ S.A.)

3.1.1. RESUMEN

En el paper “[Peña Ayala, Alejandro, Diciembre 2005]: **Inteligencia de negocios: Una propuesta para su desarrollo en las Organizaciones, Editorial APA**” se presenta y describe un modelo general OLAP y prototipo de un Sistema Data Warehouse para una empresa del sector público en general y se implementa en la empresa pública Editora Perú S.A. Se revisan los antecedentes, cómo se consolida la información actualmente de forma manual o con apoyo de otros sistemas, se define el problema, se muestra gráficamente la situación actual, se determina la justificación del presente trabajo y los métodos utilizados. Se detallan los objetivos generales y específicos; además se explica el concepto de Inteligencia de Negocios y Almacén de Datos. Se muestra el Modelo General OLAP de Editora Perú, así como el prototipo desarrollado para mostrar parte de la solución al problema.

La generación de reportes detallados, resumidos y comparativos es el medio más utilizado para explotar la información del Sistema ERP Baan, un sistema ERP (Enterprise Resource Planning), según Kwon [Kwon 2001], es “un paquete de software amplio integrado empresarial diseñado para mantener los más altos estándares de calidad de los procesos empresariales”. Para analizar la información los usuarios mensualmente generan un archivo en formato ASCII con la información de los reportes, luego mediante el procedimiento FTP (File Transfer Protocol) se importa el archivo del ambiente AIX al ambiente Windows (computador del usuario), luego desde Microsoft Office se procede a la importación y formateo del archivo, para finalmente preparar los cuadros y gráficos que serán entregados a los ejecutivos de la empresa para la toma de decisiones.

3.1.2. DEFINICION DEL PROBLEMA

Editora Perú S.A. es una empresa del sector público que ha implementado el sistema ERP Baan IV desde el año 2001, donde todo el procesamiento de transacciones en línea (OLTP) es soportado por este sistema. Como se cuenta con información histórica en el repositorio o base de datos (ORACLE), es necesario procesar, explotar, analizar, para informar a todos los niveles de decisión de la empresa. Actualmente este proceso se realiza con un uso intensivo de procesamiento en línea, el uso de herramientas diversas, el tiempo de consolidación y finalmente los resultados deben ser validados debido a que en el transcurso del proceso pueden haber variado. En la Figura 5 se muestra la situación actual, El Sistema ERP Baan se utiliza para generar reportes históricos que consume recursos en línea ocasionando lentitud y compitiendo con el trabajo diario. Cada fin de mes se solicita la generación de información consolidada, no se atiende a tiempo debido a que el procedimiento para su realización es tedioso y poco confiable.

Los problemas son los siguientes:

El uso intensivo de reportes en línea para la generación de reportes históricos compite con el procesamiento de transacciones del día, lo que ocasiona lentitud en el tiempo de respuesta del sistema.

El procesamiento de la información genera archivos ASCII, los que son formateados en Microsoft Office Excel, no garantiza la integridad debido a la manipulación de esta Información.

El ingreso de transacciones que afectan el periodo procesado ocasiona que hayan diferencias entre lo procesado y la información real.

Debido al procesamiento manual no se cuenta con la información en la fecha acordada

No hay estandarización de los cuadros y gráficos presentados como parte del análisis

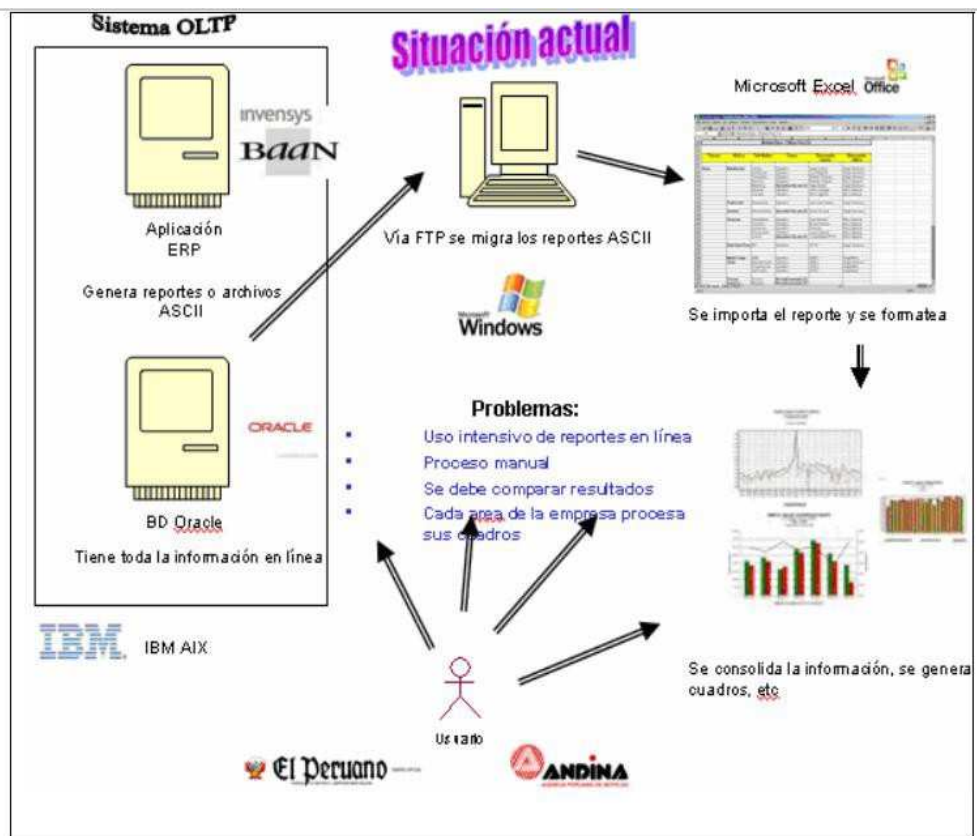


Figura 6 Situación actual de Editora Perú S.A

3.1.3. LIMITACIONES DE LA SOLUCIÓN DEL PROBLEMA

No se cuenta con partida presupuestal para el desarrollo del sistema, por lo que no se puede contratar a terceros, tampoco comprar una solución a medida, ni compra de licencias ni hardware.

Solamente se cuenta con el recurso humano del Departamento de Desarrollo para solucionar el problema.

3.1.4. VARIANTES DE LA SOLUCIÓN DEL PROBLEMA

De contar con presupuesto la solución del problema puede encargarse a un tercero para un desarrollo a medida o mediante la compra de una solución a una empresa de software

3.1.5. JUSTIFICACIÓN E IMPORTANCIA

La falta de información consolidada no permite tomar decisiones rápidas, solamente se cuenta con información mensual.

El tiempo de procesamiento de la información ocasiona que no se cumpla con la entrega oportuna.

No existe un solo repositorio de información dedicado exclusivamente a explotar la información histórica.

No existe una herramienta de análisis que permita acceder a la información histórica.

3.1.6. OBJETIVOS GENERALES

Proponer un Modelo General OLAP para Editora Perú S.A. que sirva como guía para la elaboración paulatina de los DataMarts, que permitirán atender a futuro los requerimientos de información de las diferentes áreas de la empresa.

3.1.7. OBJETIVOS ESPECIFICOS

Implementar un DataMart de Ventas a partir del Modelo OLAP SUNAT que permita solucionar el problema actual.

Utilizar la experiencia y conocimiento adquiridos en la solución del problema para la implementación de los demás DataMarts definidos en el Modelo General

3.1.8. MODELO OLAP PARA UNA EMPRESA PÚBLICA

A fin de alcanzar los objetivos planteados en el presente trabajo se ha planteado el modelamiento de una solución general para toda empresa del estado, tomamos como base la empresa pública Editora Perú, pero puede extenderse la solución a cualquier otra empresa tal como SEDAPAL, ESSALUD, Ministerios, etc.



Figura 7 Modelo OLAP general

En la Figura 7 se muestra el Modelo General en este caso mostrando a Editora Perú, pero se puede generalizar para toda empresa del estado, todas las empresas tienen la obligación legal de reportar a las siguientes entidades:

SUNAT (Superintendencia Nacional de Administración Tributaria)

FONAFE (Fondo Nacional de Financiamiento de la Actividad Empresarial del Estado)

PRODUCE (Ministerio de la Producción)

INEI (Instituto Nacional de Estadística e Informática)

Contaduría General de la República

Contraloría General de la República

CONSUCODE (Consejo Superior de Contrataciones y Adquisiciones del Estado).

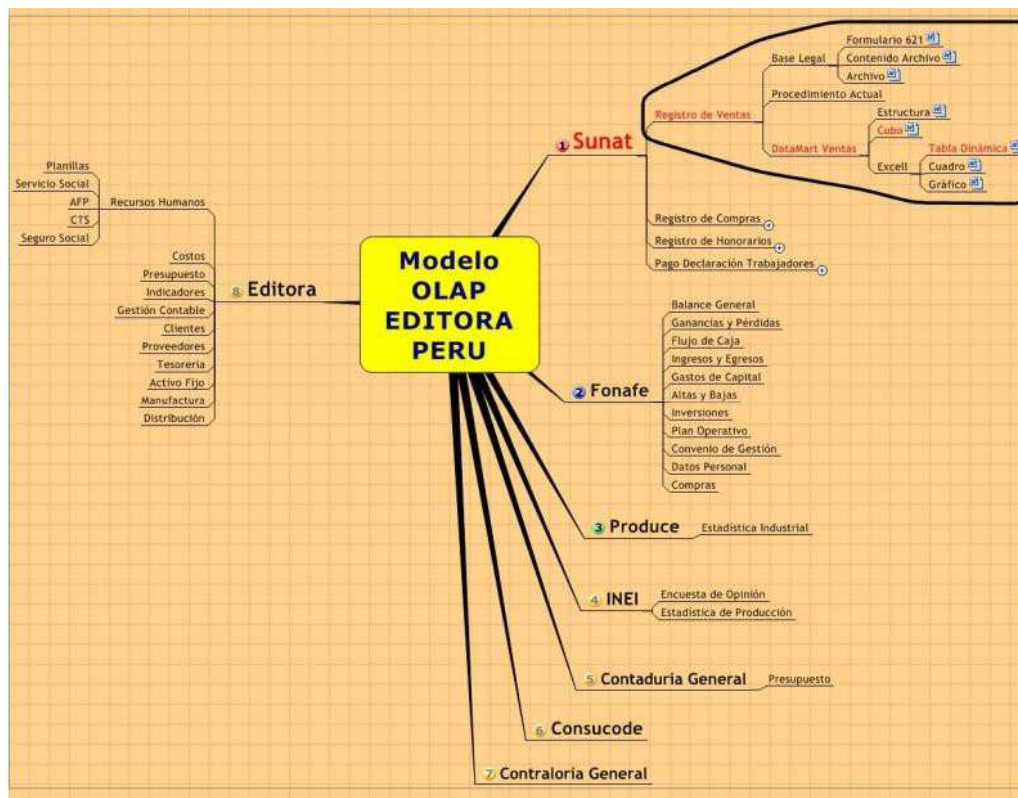


Figura 8 Modelo OLAP General detallado

En la Figura 8 podemos apreciar que en el extremo superior derecho se muestra el Modelo OLAP Sunat, se resalta el Registro de Ventas que será lo que se desarrollará como parte de este trabajo, en este caso se tiene la base legal, los formularios, formatos de archivo a informar, archivo de transferencia, luego a través del Sistema PDT la información será enviada a la Sunat.

El presente trabajo permite realizar un “reciclaje” o “reúso” de ésta información para alimentar los DataMart y conformar el Data Warehouse de una empresa pública.

3.1.9. EXPERIMENTOS Y RESULTADOS

Organización e instancias de prueba

Se ha considerado los años del 2002 al 2005. Los componentes de hardware con los que trabajará el sistema son: Servidor compatible, Servidor IBM X Series Modelo H70, SAN Storage HP.

Mientras que los componentes de software con los que trabajará el sistema son: Sistema Operativo: Windows 2000 para el servidor, Explorador de Internet (Internet Explorer), Manejador de base de datos MSSQL 2000, Sistema ERP, Sistema Operativo AIX v 4.3, Manejador de Base de Datos Oracle v.9, Sistema ERP BAAN IV C2.

Procesamiento

Se ha desarrollado procedimientos almacenados que permiten transferir los datos del DataMart de ventas de la base de datos Oracle v9.0 a Ms SQL 2000, luego otro procedimiento almacenado genera el cubo o DataMart de ventas centro del Analysis Server, en ese momento el usuario tiene disponible la información que se muestra.

Resultados

El usuario debe acceder a la hoja de cálculo y mediante un procedimiento seleccionar el cubo de ventas.

La Figura 9 nos muestra las ventas totales de los años 2002 al 2005 donde podemos visualizar cada mes como se comportan las ventas en Editora Perú.

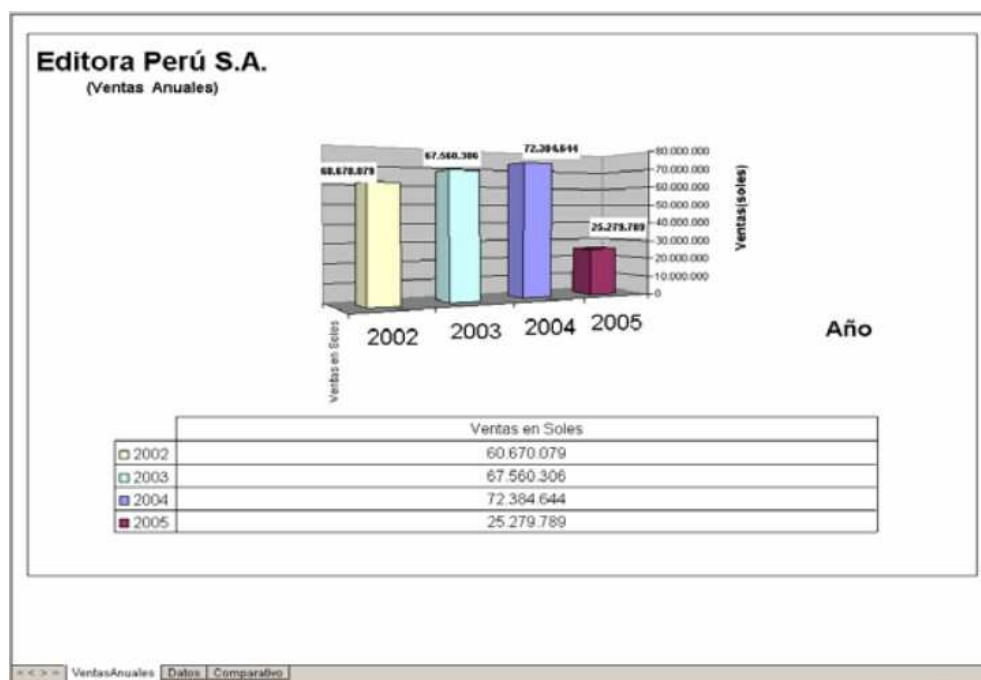


Figura 9 Ventas Anuales



Figura 10 Comparativo Mensual de ventas

Discusión de los Experimentos

En la Figura 10 se muestra gráficamente la solución al problema, así como las ventajas que se obtienen con la implementación de esta solución:

Ahora se cuenta con un servidor OLAP diferente al servidor OLTP por lo que ya no se compite con los recursos diarios, el acceso es más rápido y directamente de la hoja de cálculo Microsoft Office.

El procesamiento es totalmente automatizado, se procesa el periodo deseado, en este caso se ha considerado hacerlo quincenal y mensualmente.

Al existir un solo repositorio OLAP ahora todos los usuarios de la empresa pueden acceder simultáneamente a la información sin problemas de lentitud manteniéndose la integridad.

Se ha estandarizado el uso de plantillas para la presentación de la información tanto de los cuadros como de los gráficos.

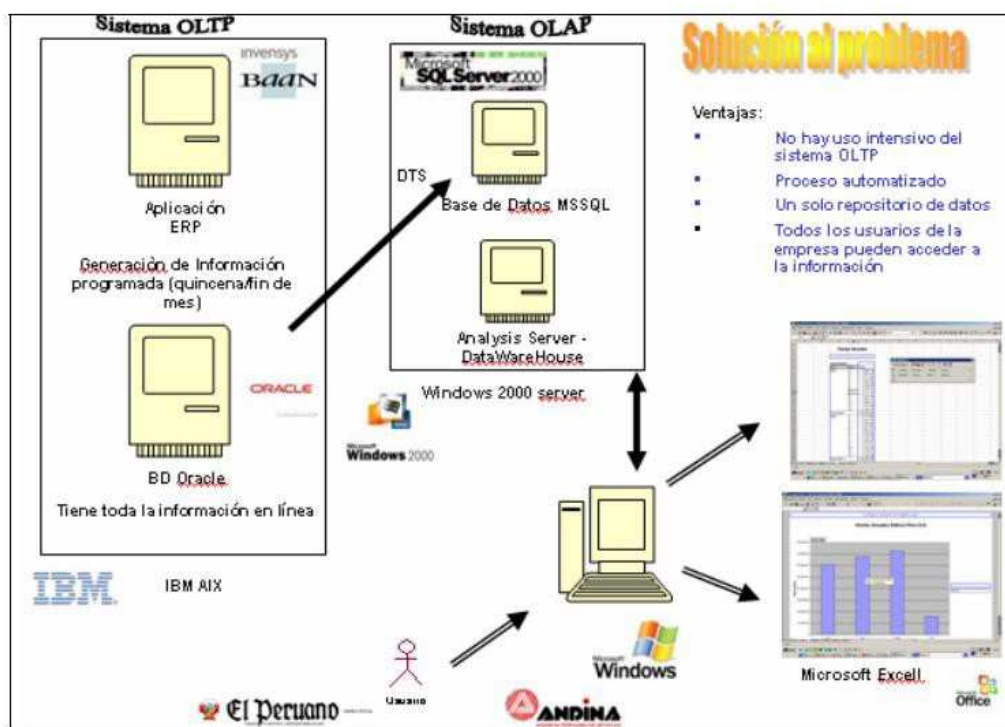


Figura 11 Solución al problema

La Figura 11 nos muestra el cuadro comparativo mensual de ventas, en él se puede comparar la venta mensual como anual, de modo que se muestren las tendencias y las expectativas de venta de acuerdo al pronóstico realizado al inicio del año. Esta información está disponible para toda la empresa.

3.1.10. CONCLUSIONES

La implementación del DataMart de ventas permitirá que el usuario pueda contar con una herramienta en línea totalmente automatizada, de fácil uso, que le permita disminuir el tiempo de procesamiento y dedique mayor tiempo a la etapa de análisis de la información.

Al automatizar la generación de reportes y gráficos:

Se eliminan los errores o diferencias por migración de datos y formateo.

Se disminuye el tiempo de procesamiento por procedimientos manuales.

Por decisión del usuario el procesamiento se realiza quincenal y mensualmente, quedando abierta la posibilidad de efectuarlo diariamente.

El acceso a la información residente ahora en el servidor OLAP es más rápido en comparación con el acceso a la base de datos del servidor OLTP utilizado en el proceso manual.

Según [Microsoft, 2004], los usuarios pueden acceder a un solo repositorio de datos (servidor OLAP) directamente desde la hoja de cálculo Microsoft Office, utilizando plantillas estándar. El único costo para la implementación de este sistema consiste en la capacitación de la herramienta Analysis Server de Microsoft producto que viene como parte de la Base de Datos MS SQL.

3.2. MODELO DE DATAWAREHOUSE PARA LA CAJA MURCIA

3.2.1. RESUMEN

Caja Murcia inauguró su primera sucursal en 1.965, desde entonces la Caja de Ahorros ha experimentado un crecimiento continuado y estable que la ha

convertido en la base del sistema bancario de la Región de Murcia y una de las principales entidades de crédito del Levante español donde en la actualidad concentra la mayor parte de sus más de 400 oficinas.

Como resultado de la favorable trayectoria de **Caja Murcia** en los últimos años, mantenida de forma constante, la Agencia Internacional Fitch Ratings ha mantenido la calificación crediticia a largo plazo de la entidad en “A+”, siendo la primera Caja de Ahorros, entre las de su dimensión, en obtener dicha calificación y superando incluso a otras de mayor tamaño.

La cohesión del equipo humano y el apoyo de los órganos de gobierno de la Caja son los verdaderos artífices de su éxito, siendo referencia en el sector como modelo de eficiencia y compromiso con sus clientes y con la sociedad en su conjunto.

3.2.2. DEFINICIÓN DEL PROBLEMA

Para obtener una valoración “sobresaliente en eficacia”, otorgada recientemente por la revista DINERO, es necesario seguir desarrollando las acciones del Plan Estratégico de CAJAMURCIA, y en concreto el Plan de Productividad. Este plan es un proyecto global de productividad en oficinas cuyo objetivo es identificar con carácter de urgencia situaciones que impiden poder atender a los clientes de forma excelente y además en operaciones que reporten beneficios.

3.2.3. OBJETIVOS GENERALES

Mejorar el tiempo de entrega de resultados de los trabajadores, y centrarlos en tareas estratégicas para el negocio.

3.2.4. PLANTEAMIENTO DE LA SOLUCION

Para obtener los resultados requeridos, era necesario romper algunas barreras e instalar elementos de autoservicio, además de llevar a cabo la implantación del modelo de productividad DataMart de Tecnocom, una potente herramienta

de gestión que permite analizar y controlar todas las operaciones y cualquier movimiento transaccional en las oficinas.

Uno de los principales medios técnicos instalados en las oficinas para conseguir el incremento de operaciones de autoservicio fue el Actualizador de Libretas. Actualmente está instalado en 135 oficinas de la Caja. Tecnocom, en su afán de llegar más allá, aprovechando la base del Actualizador, diseñó un nuevo dispositivo de entrada sobre el modelo tradicional, ampliando las posibilidades de éste a la recogida de correo y los avances de movimientos. Introduciendo la tarjeta financiera o libreta, permite solicitar la impresión en A4 de un extracto detallado de los recibos que llegaban por correo con los conceptos mucho más definidos, así como todos los movimientos de la cuenta. Esta simple operación ha desembocado también en un cambio en la manera de proceder importante: la retirada de los buzones instalados en la oficina para sustituirlos por el servicio de correo virtual, cuyo acceso es a través de una tarjeta llave que permite al cliente recoger el correo pendiente.

El seguimiento del Plan de Productividad se realiza a través del DataMart de Productividad de Tecnocom, donde de forma periódica se saca la información de todas las transacciones migrables y los porcentajes de migración obtenidos por las oficinas del Plan, haciendo una comparativa de la evolución de éstas.

Además se instaló una aplicación CRM operacional. Esta aplicación entre otras funcionalidades provee al personal de la oficina de una ficha con la situación integrada de todas las posiciones del cliente, lo que permite un mayor conocimiento de la situación para valorar sus necesidades y, por tanto, adecuar el servicio a sus preferencias.

3.2.5. RESULTADOS

Consecución de un cambio cultural entre los empleados de las oficinas.

El desencadenante de este cambio sin duda fue el modelo de productividad de Tecnocom y el elemento de confianza el Actualizador de Libreta.

El cambio tecnológico ha aumentado la productividad del personal de la Caja, liberándolos de tareas rutinarias permitiéndoles centrarse en tareas de asesoramiento y la venta de servicios de valor añadido para CAJAMURCIA, lo que al final ha repercutido en mayores ingresos.

Se ha elaborado un protocolo comercial para “cliente potencial”, una iniciativa orientada a la mejora de la atención a los nuevos clientes

3.2.6. CONCLUSIONES

Se ha aumentado en un 20% el tiempo de disponibilidad del personal de las oficinas, permitiéndoles centrarse en tareas de asesoramiento a los clientes.

3.3. MODELO DE DATAWAREHOUSE PARA EL GRUPO SANTO DOMINGO

3.3.1. RESUMEN

Con la informatización y el crecimiento de las empresas, se ha elevado en gran medida la capacidad de generación y almacenamiento de datos; son tan grandes los volúmenes de datos que su análisis no puede realizarse con los métodos tradicionales existentes. Mientras mayor es la capacidad para almacenar datos, mayor es la incapacidad para extraer información realmente útil, y gran parte de la información importante para la toma de decisiones queda oculta, y muchas veces no tomadas en cuenta debido a que no se pudo obtener esa valiosa información en el momento requerido, además los sistemas transaccionales que se utilizan comúnmente en las organizaciones no son los adecuados para análisis y extracción de información histórica y relevante. El presente trabajo se orienta a implementar un DataMart, como una herramienta que permitirá desarrollar Inteligencia de Negocios sobre el área de Marketing de una Empresa que brinda servicios de Recursos Humanos; con el presente proyecto de tesina se plantea la simplificación de los procedimientos de extracción, transformación y procesamiento de datos, para la obtención de información del comportamiento de los clientes, y con esto brindar apoyo en la toma de decisiones al área de Marketing en estudio.

3.3.2. DEFINICION DEL PROBLEMA

Debido a que no existe una fuente de información centralizada, integrada y homogeneizada, que permita apoyar al área de Marketing en la toma de decisiones de forma oportuna y relevante, ya que la integración y homogeneización de los datos que son relevantes para la toma de decisiones del área de Marketing de la empresa no están bien definidas.

3.3.3. OBJETIVOS

Apoyar a la de toma de decisiones del área de marketing a través de la entrega oportuna de información mediante la construcción de un DataMart.

Definir mecanismos que permitan automatizar la recolección y homogeneización de los datos.

Definir los datos y las fuentes de datos que son relevantes para el área de marketing y la toma de decisiones con respecto a los clientes y servicios de la empresa.

3.3.4. PROPUESTA

Para la solución del problema definido en el presente proyecto, se plantea la implementación de un DataMart, que implica lograr la transformación, homogeneización e integración los datos, para luego obtener reportes que permitan satisfacer la necesidad de información del área de Marketing, y a su vez todo esto les brinde apoyo y una base para la toma de decisiones

3.4. POSIBLES HERRAMIENTAS A UTILIZAR

3.4.1. SSIS

El software SQL Server Integration Services (SSIS) [SQL 2007], permite la integración de los datos de cualquier fuente. SSIS provee una plataforma escalable y extendible que capacita al equipo desarrollador a construir, mantener, y desplegar soluciones de integración para alcanzar soluciones de

integración únicas de acuerdo a las necesidades. Destacan sus herramientas de minería de datos y administración de objetos.

3.4.2. DATA STAGE

DataStage [DAT 2007] es una herramienta que permite soportar la información que necesita la compañía, y construir un Data Warehouse en “tiempo real”. El DataStage es una herramienta ETL (Extract/Transform/Load - Extracción, Transformación y Carga) que utiliza notación gráfica para construir integración de datos para dar soluciones, y está disponible en varias versiones, como Server Edition y Enterprise Edition. Es una de las herramientas ETL más rápidas y potentes del mercado.

3.4.3. SUNOPSIS

También existe en el mercado, Sunopsis [SUN 2007], que ofrece un alto desempeño y una integración efectiva, cubriendo las necesidades de integración. Esta herramienta permite el desarrollo y el mantenimiento simple, que permite que los proyectos de integración se realicen a tiempo y en presupuesto. Sinopsis trabaja con una arquitectura ELT (Extracción, Load, Transform) en lugar de la tradicional ETL.

3.4.4. MICROSTRATEGY

Existen soluciones como MicroStrategy Business Intelligence Solutions que permite mejorar y predecir el comportamiento del negocio, poniendo información en las manos de toda persona de negocios en la empresa. Esta tecnología ofrece capacidades de monitoreo, de reportes y de análisis, que permiten tomar mejores decisiones cada día, y lograr las metas planteadas en cada organización. Esta herramienta permite la generación de scorecards y dashboards, reportes, análisis OLAP, análisis avanzado y predictivo, alertas y notificaciones.

3.4.5. COGNOS

Cognos 8 Business Intelligence [COG 2007] es una plataforma del grupo IBM que permite la generación y visualización de reportes, cubos, dashboards y Balance scorecards, además de la gestión de permisos y usuarios necesaria para la implementación de la plataforma

3.4.6. BUSSINESS OBJECTS

Bussiness Objects es otra plataforma BI que se caracteriza por ofrecer distintas funcionalidades según el tamaño de la empresa que la adquiere y la licencia.

3.4.7. PENTAHO

Pentaho es una plataforma de BI “orientada a la solución” y “centrada en procesos” que incluye todos los principales componentes requeridos para implementar soluciones basados en procesos. Las soluciones que Pentaho pretende ofrecer se componen fundamentalmente de una infraestructura de herramientas de análisis e informes integrados con un motor de workflow de procesos de negocio. La plataforma será capaz de ejecutar las reglas de negocio necesarias, expresadas en forma de procesos y actividades y de presentar y entregar la información adecuada en el momento adecuado. [28]

3.4.8. OCTOPUS

Octopus es, al igual que Pentaho, una herramienta libre pero sólo se centra en los procesos ETL. Está basada en Java y por lo tanto se puede conectar a cualquier fuente JDBC.

3.5. CUADRO COMPARATIVO DE LAS HERRAMIENTAS

3.5.1. DEFINICION DE LAS VARIABLES DE COMPARACION

En la Tabla 3.1 mostramos la definición y ponderación de las variables de comparación para las herramientas antes mencionadas.

Variables	Ponderación	Definición
ETL	1	Permite la Extracción, Transformación y Carga.
Elaboración de Reportes	1	Permite la elaboración de reportes
Open Source	1	El código fuente de la solución es de código abierto
Conocimiento herramienta	2	Conocimiento de la herramienta tiene una ponderación de 2 porque consideramos una ventaja el manejo de la herramienta para realizar el proyecto

Tabla 3.1 Definición y ponderación de variables de Comparación de herramientas BI

Fuente: Elaboración Propia

3.5.2. CRITERIO DE PUNTUACION

En la tabla 3.2 se muestra una puntuación que se asigna al valor de las variables

Criterios de puntuación	puntaje
SI	1
NO	0

Tabla 3.2 Cuadro de puntuación de las herramientas BI

Fuente: Elaboración Propia

3.5.3. PUNTUACION DE LAS HERRAMIENTAS

En la tabla 3.3 se aprecia que SQL Server Integration Services obtiene el mayor puntaje. En la Figura 12 se puede apreciar el puntaje en porcentajes siendo SSIS el que obtiene el 100% de puntuación.

Tomando como base las características de estas herramientas y el puntaje obtenido se optó por utilizar la herramienta SSIS(SQL Server IntegrationServices)

Herramienta	ETL	Elaboración de Reportes	Open Source	Conocimiento herramienta	Puntuación
SSIS	1	1	0	2	4
Data Stage	1	0	0	0	1
Sunopsis	1	0	0	0	1
Micro Strategy	0	1	0	0	1
BussinessObjects	0	1	0	0	1
Cognos	0	1	0	0	1
Octopus	1	0	0	0	1
Pentaho	1	1	1	0	3

Tabla 3.3 Cuadro comparativo de herramientas BI

Fuente: Elaboración Propia

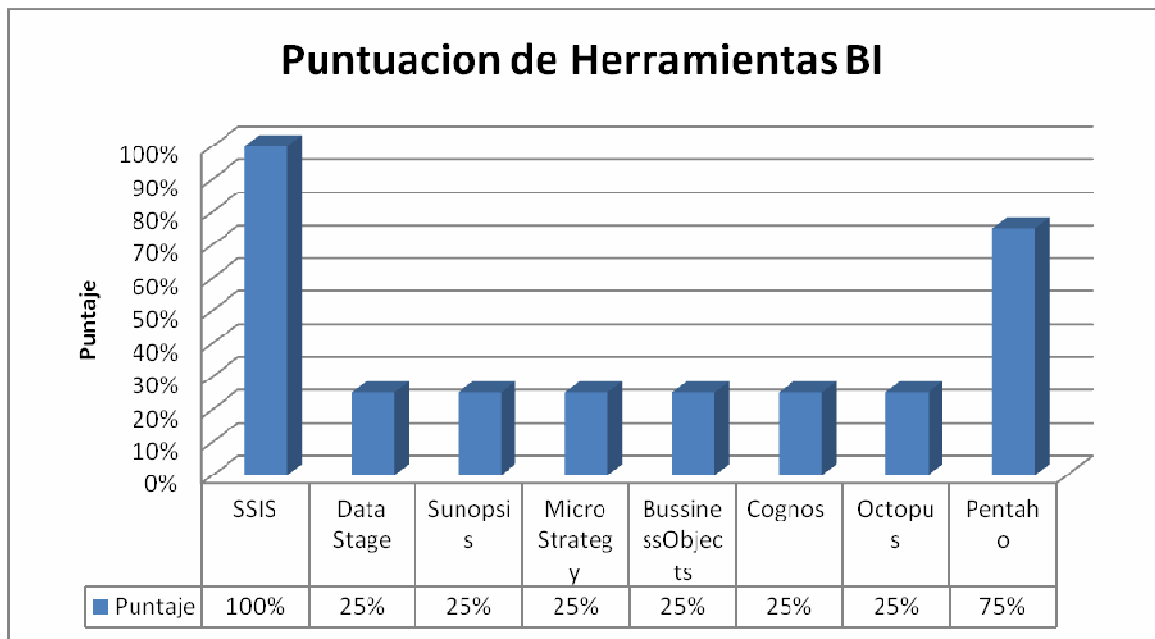


Figura 12 Cuadro comparativo de herramientas de inteligencia de negocios

Fuente: Elaboración Propia

3.6. POSIBLES METODOLOGIAS A UTILIZAR

3.6.1. METODOLOGIA ROADMAP

En términos generales de BI, el Business Intelligence RoadMap (BIR) especifica el camino y la dirección que deben seguir las aplicaciones, estructuras, herramientas y personas que intervienen en un proyecto de este tipo. Este BIR es, en primer lugar, una guía del ciclo de vida de un proyecto para desarrollar aplicaciones de soporte a la decisión utilizando datos estructurados. El 60% de los proyecto de BI son abandonados o acaban fallando por una inadecuada planificación, por falta de tareas, por entregas fuera de plazo, por una mala gestión del proyecto, por una ausencia de requisitos de negocio o por una mala calidad en las entregas. Los gestores necesitan saber qué hacer y no hacer en implementaciones de BI, basándose en experiencias fiables. [26]

En el BIR se definen las actividades necesarias en un proyecto BI para mantener la integración de la infraestructura del entorno al que pertenecen. Las infraestructuras técnicas y no técnicas son las competencias esenciales para la alineación organizacional. Además, se deben definir los roles y responsabilidades asignados a cada persona del equipo para cada paso del desarrollo. BIR describe 16 pasos a seguir en un proyecto de BI, los cuales se encuentran repartidos dentro de los seis estados que se han descrito antes como necesarios para crear un proyecto de ingeniería con éxito. A continuación se van a describir cada uno de ellos dentro de su estado correspondiente. [26]

3.6.2. METODOLOGÍA DE KIMBALL

Se enfoca principalmente en el diseño de la base de datos que almacenará la información para la toma de decisiones. El diseño se basa en la creación de tablas de hechos (FACTS) que son tablas que contienen la información numérica de los indicadores a analizar, es decir la parte cuantitativa de la información.

“Ralph Kimball, es reconocido como uno de los padres del concepto de Data Warehouse, se ha dedicado desde hace más de 10 años al desarrollo de su metodología para que éste concepto sea bien aplicado en las organizaciones y se asegure la calidad en el desarrollo de estos proyectos.”

3.6.3. METODOLOGIA DATAWAREHOUSE ENGINEERING PROCESS (DWEP)

Metodología basada en RUP y en la herramienta UML para desarrollar un Data Warehouse o DataMart llamada Data Warehouse Engineering Process (DWEP). Esta metodología fue propuesta por Sergio Luján-Mora y Juan Trujillo en el año 2006. Para el presente proyecto dicha metodología será aplicada al desarrollo de un DataMart.

El DWEP divide el desarrollo del almacén de datos, al igual que en RUP, en 4 fases: Concepción, Elaboración, Construcción y Transición. Los objetivos generales por cada fase son los mismos que en RUP.

3.7. CUADRO COMPARATIVO DE METODOLOGIAS

3.7.1. DEFINICION DE LAS VARIABLES DE COMPARACION

En la Tabla 3.4 mostramos la definición y ponderación de las variables de comparación para las metodologías antes mencionadas.

Variables	Ponderación	Definición
Metodología genérica	1	Aplicable a múltiples plataformas.
Ciclo de vida	1	Considera todo el ciclo de vida del producto, hasta el despliegue y la marcha.
Planeación estratégica	1	Cubre la planeación estratégica de la organización.
Etapas detalladas	1	Tiene etapas completamente detalladas, que sumándolas dan como resultado la completitud del proyecto de BI.
Actividades de cada etapa	1	Cada etapa contiene las actividades a realizar para lograr su completitud.

Tabla 3.4 Definición y ponderación de variables de Comparación de las metodologías

Fuente: Elaboración Propia

3.7.2. CRITERIO DE PUNTUACION

En la tabla 3.5 se muestra una puntuación que se asigna al valor de las variables

Criterios de puntuación	puntaje
SI	1
NO	0

Tabla 3.5 Cuadro de puntuación de las metodologías BI

Fuente: Elaboración Propia

3.7.3. PUNTUACION DE LAS METODOLOGIAS

En la tabla 3.6 se aprecia que la metodología ROADMAP obtiene el mayor puntaje. En la Figura 13 se puede apreciar el puntaje en porcentajes siendo ROADMAP el que obtiene el 100% de puntuación.

Tomando como base las características de estas metodologías y el puntaje obtenido se optó por utilizar la metodología ROADMAP

Herramienta	Metodología genérica	Ciclo de vida	Planeación estratégica	Etapas detalladas	Actividades de cada etapa	Puntuación
ROADMAP	1	1	1	1	1	5
KIMBALL	1	1	0	1	1	4
DWEP	1	0	0	1	0	2

Tabla 3.6 Cuadro comparativo de metodologías BI

Fuente: Elaboración Propia

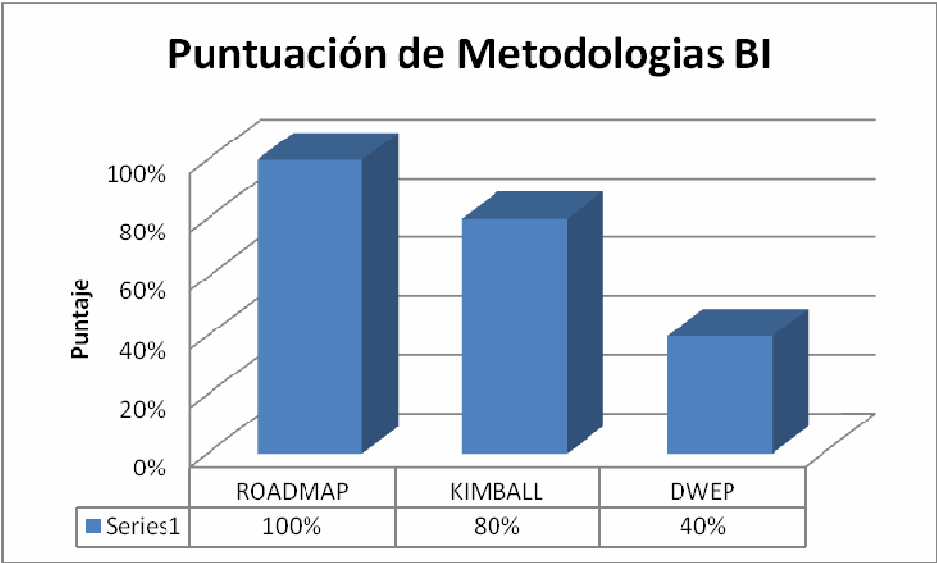


Figura 13 Cuadro comparativo de metodologías de inteligencia de negocios

Fuente: Elaboración Propia

CAPITULO IV. IMPLEMENTACIÓN Y RESULTADOS

4.1. DEFINICION DE LAS HERRAMIENTAS A UTILIZAR

A continuación mostramos las herramientas que nos ayudaran a implementar la solución de inteligencia de negocios:

4.1.1. MICROSOFT SQL SERVER 2005 ANALYSIS SERVICES (SSAS)

Proporciona la base de una solución de Inteligencia de Negocios.

Facilita la creación de sofisticadas soluciones de procesamiento analítico en línea (OLAP) y minería de datos.

Proporcionan la capacidad de diseñar, crear y administrar cubos y modelos de minería de datos de los almacenes de datos, y permiten que el cliente pueda obtener acceso a los datos OLAP y de la minería de datos.

4.1.2. MICROSOFT SQL SERVER 2005 INTEGRATION SERVICES (SSIS)

Facilita la creación de soluciones complejas y sólidas para la extracción, transformación y carga (ETL) de datos.

Proporcionan la capacidad de diseñar, crear, implementar y administrar paquetes que tratan los requisitos empresariales cotidianos.

4.1.3. MICROSOFT SQL SERVER 2005 REPORTING SERVICES (SSRS)

Es un servidor de reportes.

4.2. DEFINICION DE LA METODOLOGIA ELEGIDA

Para el desarrollo del DataMart como solución de Inteligencia de negocios para T-IMPULSO, se empleará la guía Business Intelligence RoadMap elaborado por Larissa T. Moss y ShakuAtre [26], Y la guía Metodologías de Soluciones Cognos y Cognos Business Intelligence RoadMap, dado que brinda los pasos a seguir para el ciclo de vida de un proyecto de inteligencia de negocios.

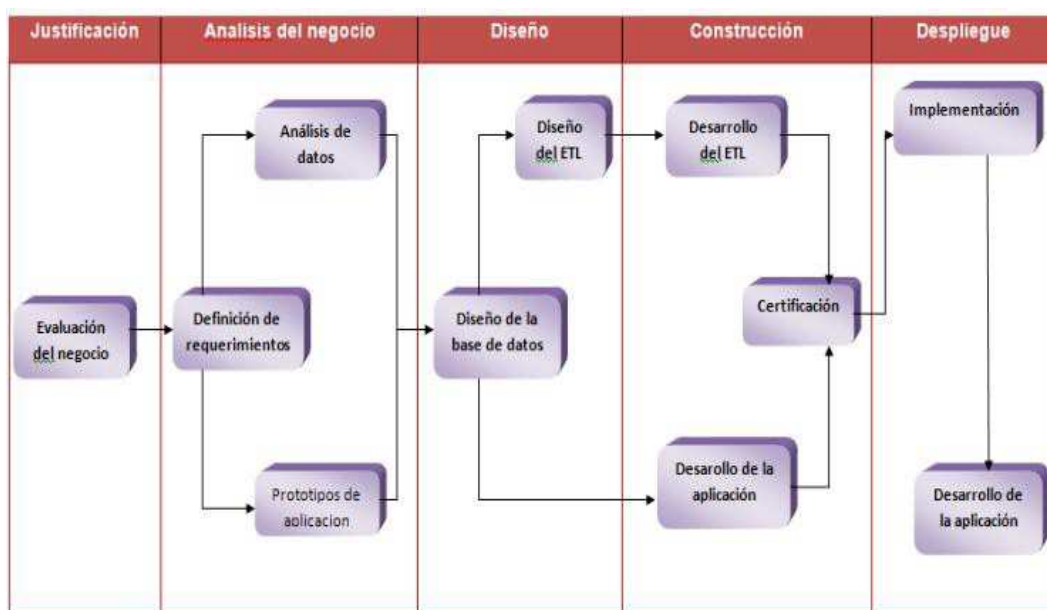


Figura 14 Metodología adaptada de Business Intelligence RoadMap

Fuente: Elaboración propia

4.2.1. PASOS PARA LA IMPLEMENTACION DEL DATAMART COMO SOLUCION DE INTELIGENCIA DE NEGOCIOS SEGÚN BUSINESS INTELLIGENCE ROADMAP

La presente tesina solo está contemplada hasta el paso “Desarrollo del ETL”, Los pasos siguientes dentro del Business Intelligence RoadMap (Desarrollo de la Aplicación, Certificación E Implementación) no son incluidos dentro del presente proyecto, puesto que estos pasos son definidos para la puesta en producción de la solución desarrollada, y el alcance de la presente tesina solo llega hasta la obtención de la solución y algunas pruebas Alfa con la misma.

1. Evaluación

Estado del acceso de la información en la empresa.

Reglas del negocio.

Estado del entorno de soporte a la toma de decisiones.

Justificación de decisiones.

Retorno de inversión.

2. Definición de Requerimientos

Requerimientos funcionales.

Requerimientos de información.

Requerimientos de seguridad y performance.

3. Análisis de Datos

Fuente de datos.

Calidad de datos.

Limpieza de datos.

4. Prototipo de Aplicación

Objetivos de prototipos.

Participación de stakeholders.

Herramientas y métodos para el prototipo.

5. Diseño de la Base de Datos

Demanda de reportes y consultas.

Consideraciones para el diseño de la base de datos.

Selección de un DBMS.

6. Diseño del ETL

Selección de herramientas ETL.

Fases de ETL.

Flujo de procesos ETL.

Métricas de calidad.

7. Desarrollo del ETL

Extracción desde sistemas fuente.

Herramientas ETL.

Flujo del desarrollo ETL.

Pruebas.

8. Desarrollo de la Aplicación

Resultados del prototipo.

Herramientas de acceso y análisis de la información.

Capacitación de usuarios.

Alcance de la aplicación.

9. Certificación

Elaboración de casos de prueba.

Pruebas de carga.

Pruebas de funcionalidades.

10. Implementación

Componentes para producción.

Mantenimiento de la base de datos.

Capacitación y soporte a la solución.

4.3. APLICACIÓN DE LA GUIA BUSINESS INTELLIGENCE ROADMAP

La presente tesina solo está contemplada hasta el paso DESARROLLO DEL ETL, Los pasos siguientes dentro del Business Intelligence RoadMap (Desarrollo de la Aplicación, Certificación E Implementación) no son incluidos dentro del presente proyecto, puesto que estos pasos son definidos para la puesta en producción de la solución desarrollada, y el alcance de la presente tesina solo llega hasta la obtención de la solución y algunas pruebas Alfa con la misma.

4.3.1. EVALUACIÓN DEL NEGOCIO

4.3.1.1. ESTADO DEL ACCESO DE LA INFORMACIÓN EN LA EMPRESA

Actualmente la empresa T-Impulso cuenta con una base de datos, la cual es utilizada por los sistemas transaccionales de la organización; entre los sistemas que utilizan la base de datos tenemos:

Sistema de Compras.

Logística.

Contabilidad.

Recursos humanos

Tesorería

Bienestar Social

Para la transferencia de datos se utilizan aplicaciones:

La intranet de la organización

Sistemas web para permitirse operar en diferentes áreas geográficas.

ERP Propios de la Empresa

4.3.1.2. REGLAS DEL NEGOCIO

Las reglas de negocio implícitas dentro de la toma de decisiones del área de Logística son:

Todo requerimiento de información que no puede ser cubierto por la misma área de Logística, se delega la responsabilidad al área de informática para que brinde las facilidades necesarias.

Todo requerimiento de información del área de Logística, debe ser evaluado y autorizado por el jefe del área.

Los requerimientos de información tienen un plazo de entrega con el mutuo acuerdo entre el área de Logística y el área de informática.

4.3.1.3. ESTADO DEL ENTORNO DE SOPORTE A LA TOMA DE DECISIONES

En la Actualidad el soporte a la toma de decisiones en el área de Logística de la empresa T-Impulso se realiza de la siguiente forma:

- a. Los trabajadores del área de logística si tienen un requerimiento de información llenan su solicitud de de requerimiento de información dirigido al área de informática de la empresa.
- b. El jefe de área de Logística aprueba la solicitud de requerimiento de información, y posteriormente lo dirigen a mesa de partes.

- c. La solicitud de Requerimiento pasa a mesa de partes, quienes canalizan el requerimiento al área de informática de T-Impulso.
- d. En el área de informática de T-Impulso, el jefe del área designa a un analista programador quien resolverá el requerimiento.
- e. El analista programador retorna el requerimiento hacia el solicitante del área de Logística, a través de un reporte, hoja de cálculo o gráfico.
- f. El trabajador del área de Logística quien solicitó el requerimiento verifica si se cumplió con el requerimiento y luego da su conformidad.

4.3.2. DEFINICIÓN DE REQUERIMIENTOS

4.3.2.1. REQUERIMIENTOS FUNCIONALES

La identificación de los requerimientos es una etapa importante para el desarrollo de este proyecto de tesis ya que dependerá en gran medida del alcance y el desarrollo del DataMart para el Área de Logística propuesta. El trabajo realizado se ha basado en entrevistas de levantamiento de información en el Área de Logística, en donde se cubrieron los siguientes puntos:

Gestión de la información en el Área de Logística.

Identificación de objetivos dentro del Área.

Identificación de los lineamientos para alcanzar la visión de la Organización.

Identificación de las lógicas del negocio.

Definición de métricas.

Publicación y distribución de la información.

Los puntos mencionados dieron lugar a una serie de requerimientos funcionales el cual detallamos a continuación.

Los requerimientos funcionales cubren: los objetivos del área, la lógica de negocio y las métricas necesarias para realizar el análisis requerido dentro del área de Logística.

Los Objetivos del área en cuanto a la función que debe cumplir la información se pudieron recopilar a través de reuniones en las cuales se entrevistaron a las siguientes personas:

Encargado de la base de datos de Logística.

Jefe de análisis cuantitativo de Logística.

Gerente del Área de inteligencia de negocios para el área de Logística

Los objetivos identificados fueron:

Adquisición de nuevos clientes.

Eficiencia en la captación de clientes de calidad.

Eficiencia en recopilación y actualización de la información de clientes.

Recuperar clientes.

Retención de los productos.

Retención de clientes

4.3.2.2. REQUERIMIENTOS DE INFORMACIÓN

Una de las necesidades de información de área de Logística es conocer las ventas realizadas por la empresa T-Impulso, en cualquier periodo de tiempo, de cualquier producto o servicios y sea el área geográfica que el usuario lo requiera.

4.3.2.3. REQUERIMIENTOS DE SEGURIDAD Y PERFORMANCE

Solo los usuarios del área de Logística deben tener acceso al DataMart, para poder realizar la extracción de la información que requieran.

4.3.3. ANÁLISIS DE DATOS

4.3.3.1. FUENTE DE DATOS

El área de Logística de la empresa T-Impulso, trabaja con 5 bases de datos en los cuales se guarda la información de los clientes, ventas y productos. Las bases de datos son de diferentes proveedores, y estos son:

Oracle 10g
SQL Server de Microsoft
MySQL de Oracle.
MySql
IbExpert

4.3.3.2. CALIDAD DE DATOS

Las tablas no se encuentran muy bien normalizadas, y además la documentación que explica la creación y los cambios realizados en las bases de datos no es suficiente como para garantizar que los datos a utilizar son los correctos.

Debido a que los datos se encuentran en RDBMS de diferentes proveedores de software, existe incompatibilidad de formatos a la hora de relacionar la data de todas las fuentes; con esto podemos decir que la calidad de los datos es mala a regular.

4.3.3.3. LIMPIEZA DE DATOS

En esta fase se comenzó a definir todas aquellas tablas (clientes, compras, productos y ubicación geográfica) en todas las RDBs que encuentran relación con las dimensiones especificadas en nuestros modelos físicos y lógicos.

4.3.4. PROTOTIPO DE APLICACIÓN

4.3.4.1. OBJETIVOS DE PROTOTIPOS

Los Objetivos del prototipo son:

Permitir realizar la carga, transformación y extracción de los datos.

Mostrar la información en forma que sea entendible al usuario.

4.3.4.2. PARTICIPACIÓN DE STAKEHOLDERS

Los stakeholders identificados son:

Encargado de la base de datos de Logística.

Jefe de análisis cuantitativo de Logística.

Gerente del Área de inteligencia de negocios para Logística

Estos brindan todas las facilidades para la realización del proyecto.

4.3.5. DISEÑO DE LA BASE DE DATOS

4.3.5.1. ESQUEMA DE LA BD ACTUAL DE LOGISTICA

Modelo de Datos físico

A continuación se presenta los modelo de datos físico para el Data Mart que representa los diferentes requerimientos que se atienden normalmente, Existen tres procesos en el área de Logística de los cuales mencionaremos dos.

Modelo físico del proceso de cotización

El proceso empieza cuando se requiere atender un requerimiento y no hay productos, se genera una nota de pedido para realizar un cotización, de acuerdo a la cotización más barata, es que se genera una orden de compra la cual se adjunta al comprobante, la factura de pago del proveedor, después se alimentó el stock del almacén y se atiende el requerimiento con una nota de salida, en la figura 15 se observa el diseño del proceso.

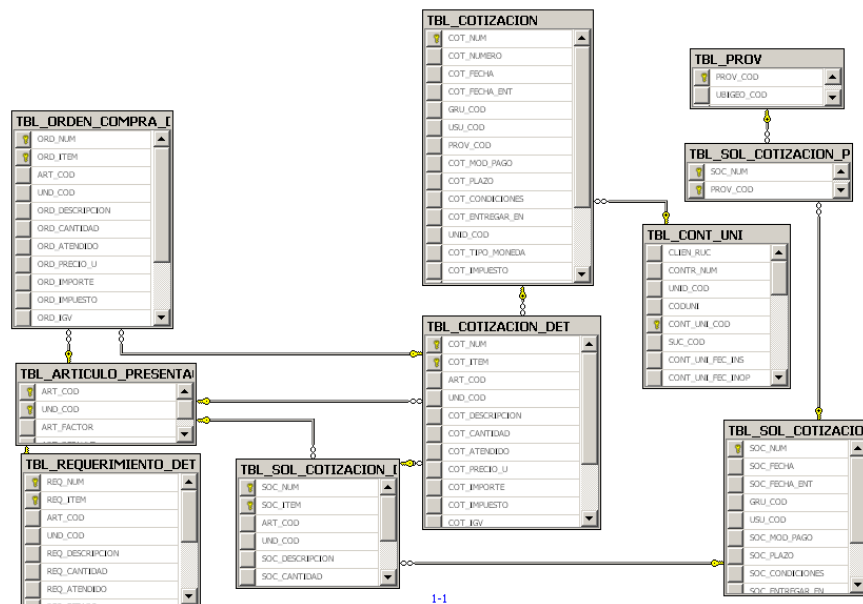
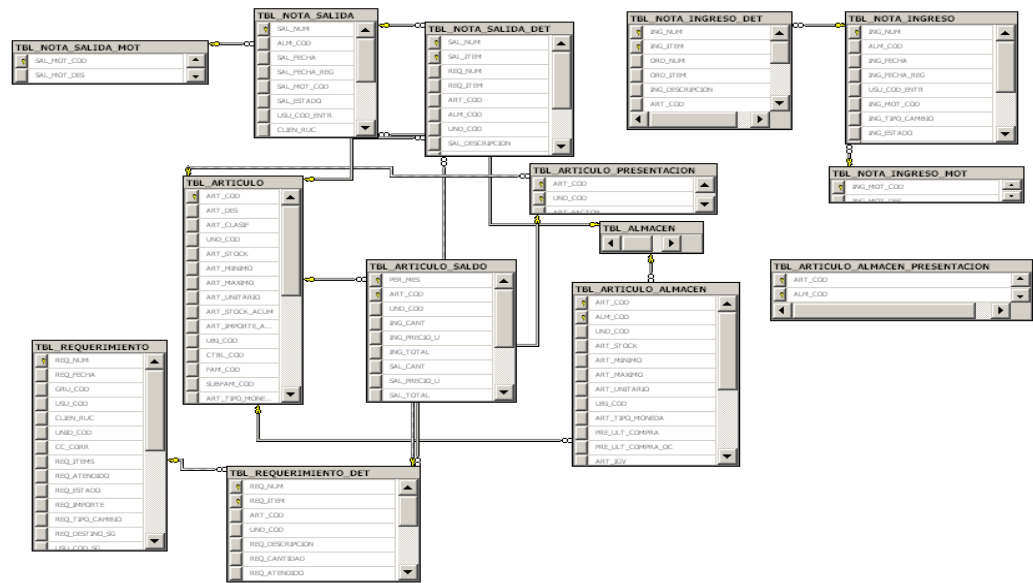


Figura 15 Modelo Físico del proceso de Cotización

Fuente: Elaboración Propia

4.3.5.2. PROCESO DE ENTRADA Y SALIDA DE PRODUCTOS

En este proceso se observa la alimentación del stock del almacén, a través de la nota de entrada, y a través de la nota de salida se atienden los requerimientos y disminuye el stock, en la figura 16 se observa el proceso de lo mencionado.



1-1

Figura 16 Modelo Físico del proceso de Entrada y Salida

Fuente: Elaboración propia

4.3.5.3. DEMANDA DE REPORTES Y CONSULTAS

Las principales demandas de reportes y consultas que tienen los trabajadores del área de Logística, son reportes de compras de acuerdo a los siguientes criterios:

- Calendario.
- Sucursal.
- Usuario.
- Proveedor
- Compras
- Orden de Compra
- Artículo

4.3.5.4. CONSIDERACIONES PARA EL DISEÑO DEL DATAMART

En la figura 17 se Construye la vista del cubo, aquí es donde se escogen solo las tablas que se utilizaran en la implantación del cubo o DataMart

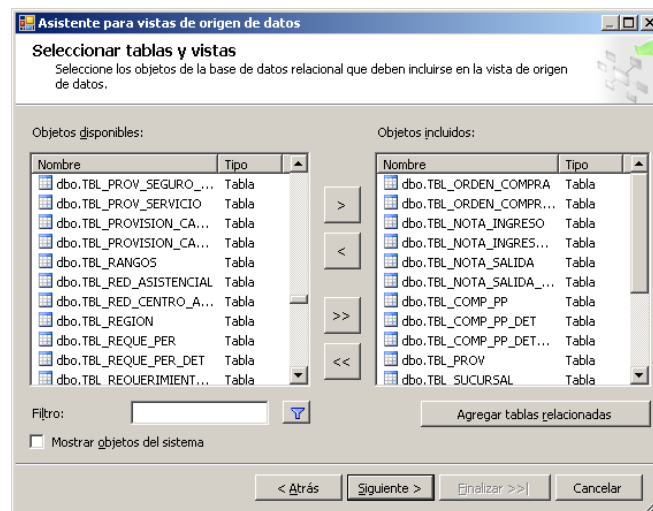


Figura 17 Creación del Origen de Datos para el DataMart

Fuente: Elaboración propia

Una vez seleccionada las tablas a utilizar se procederá a crear las vistas tal y como se observa en la figura 18.

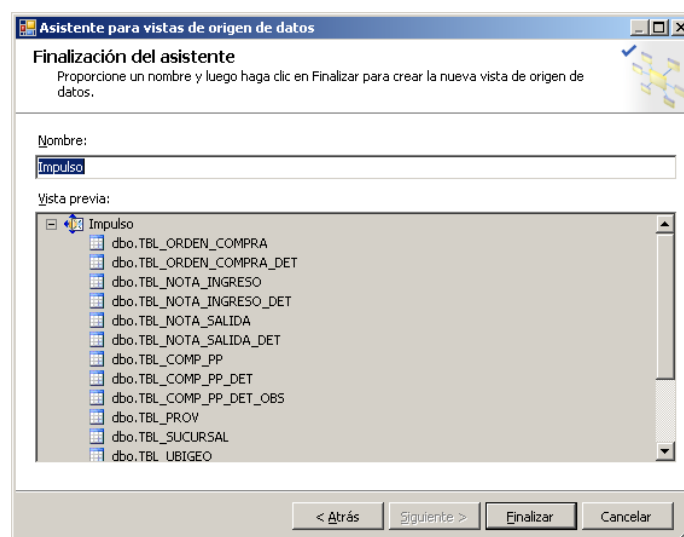


Figura 18 Selección de tablas para la creación del Origen de Datos

Fuente: elaboración propia

Una vez seleccionadas las tablas que crearan las vistas obtenemos el bosquejo de nuestra vista de Datos para el cubo.

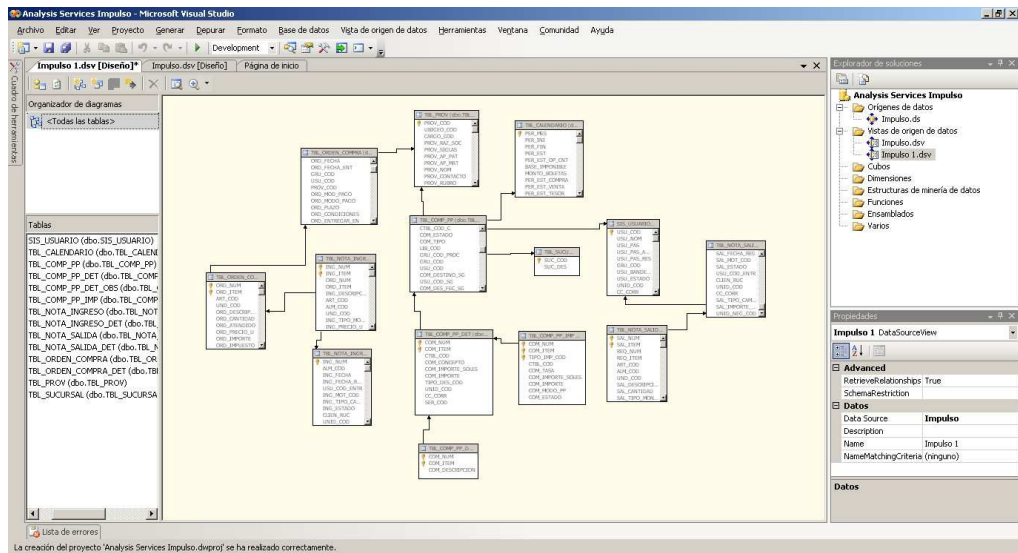


Figura 19 Vista del Origen de Datos

Fuente: Elaboración propia

Figura 20 Creación del Cubo con el Origen de Datos

Fuente: Elaboración propia

Se procede a seleccionar la vista de datos creada, para proceder a implementar el cubo

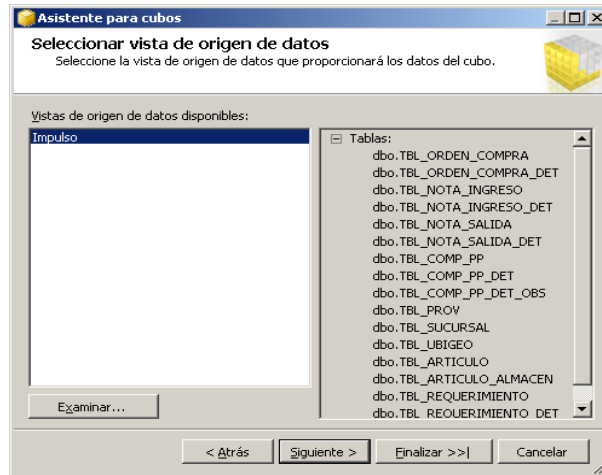


Figura 21 Creación del Cubo con el Origen de Datos

Fuente: Elaboración propia

Se procede a crear la tabla de hecho con sus dimensiones



Figura 22 Creación de la tabla de hechos con sus dimensiones

Fuente: Elaboración propia

Se procede a seleccionar de todas las tablas creadas cual será nuestra tabla de Hechos y cuáles serán las tablas de dimensiones asociadas a la tabla de hechos

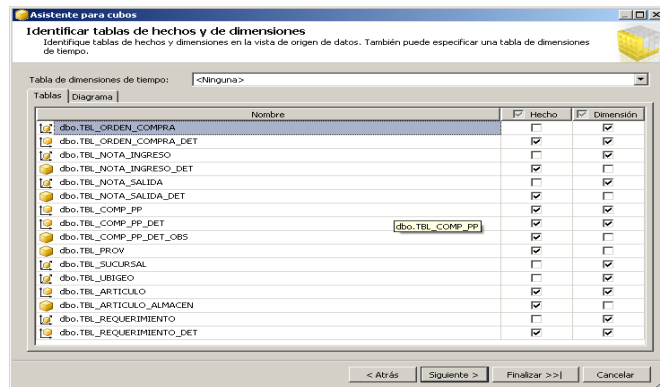


Figura 23 Selección de tabla de hechos y sus dimensiones

Fuente: Elaboración propia

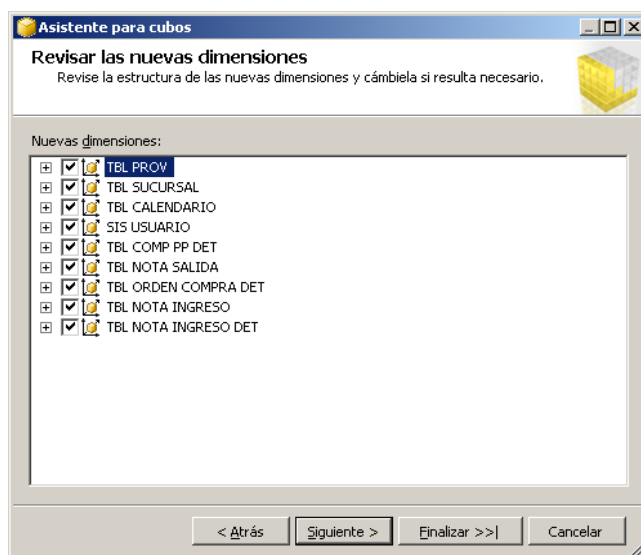


Figura 24 Creación de Dimensiones del Cubo

Fuente: Elaboración propia

En la figura 25 observamos los cubos creados en la parte superior y sus dimensiones en la parte inferior.

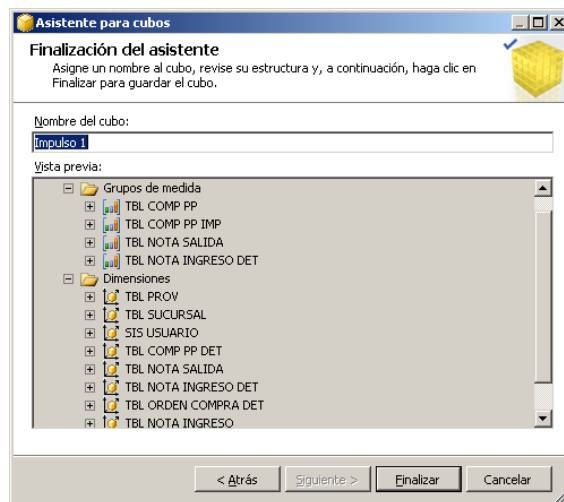


Figura 25 Creación del Cubo con sus Dimensiones

Fuente: Elaboración propia

En la figura 26 se observa el cubo creado y a partir de este cubo se proceden a realizar la elaboración de los reportes, Aquí observamos varios cubos, pero solo nos enfocaremos en el Cubo respecto a las compras

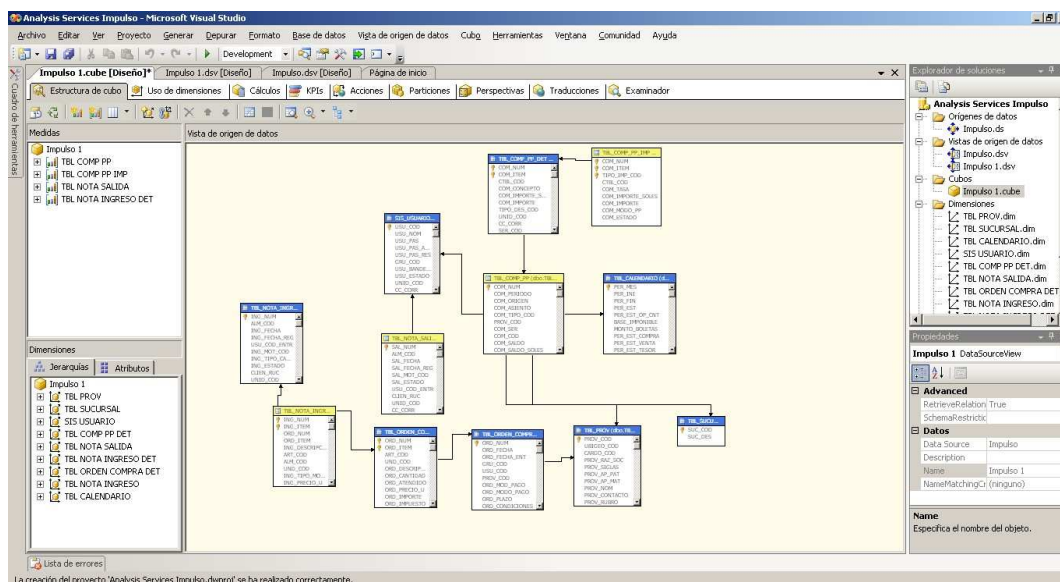


Figura 26 Creación del Cubo

Fuente: Elaboración propia

4.3.5.5. SELECCIÓN DE UN DBMS.

A continuación mostramos el DBMS que nos ayudara a implementar la solución de inteligencia de negocios para la empresa T-Impulso

Microsoft SQL Server 2005 Analysis Services (SSAS)

Proporciona la base de una solución de Inteligencia de Negocios.

Facilita la creación de sofisticadas soluciones de procesamiento analítico en línea (OLAP) y minería de datos.

Proporcionan la capacidad de diseñar, crear y administrar cubos y modelos de minería de datos de los almacenes de datos, y permiten que el cliente pueda obtener acceso a los datos OLAP y de la minería de datos.

Microsoft SQL Server 2005 Reporting Services (SSRS)

Es un servidor de reportes, que brinda la flexibilidad necesaria para satisfacer los requerimientos de información de los usuarios del DataMart.

4.3.6. DISEÑO DEL ETL

4.3.6.1. SELECCIÓN DE HERRAMIENTAS ETL

Microsoft SQL Server 2005 Integration Services (SSIS)

Facilita la creación de soluciones complejas y sólidas para la extracción, transformación y carga (ETL) de datos.

Proporcionan la capacidad de diseñar, crear, implementar y administrar paquetes que tratan los requisitos empresariales cotidianos.

4.3.6.2. FASES DE ETL

En el gráfico 27 mostramos el diseño de extracción y transformación de carga de la información necesaria, para poder después manipular los datos y trabajarlos en nuestro DataMart.



Figura 27 Proceso de Extracción, transformación y Carga

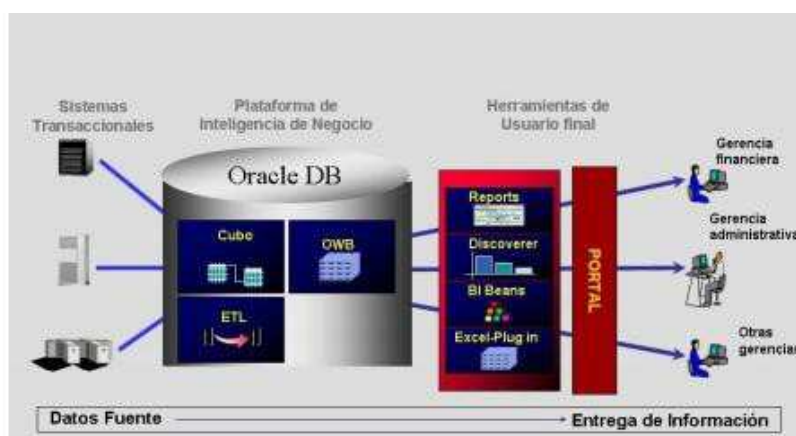


Figura 28 Plataforma de Inteligencia de negocios

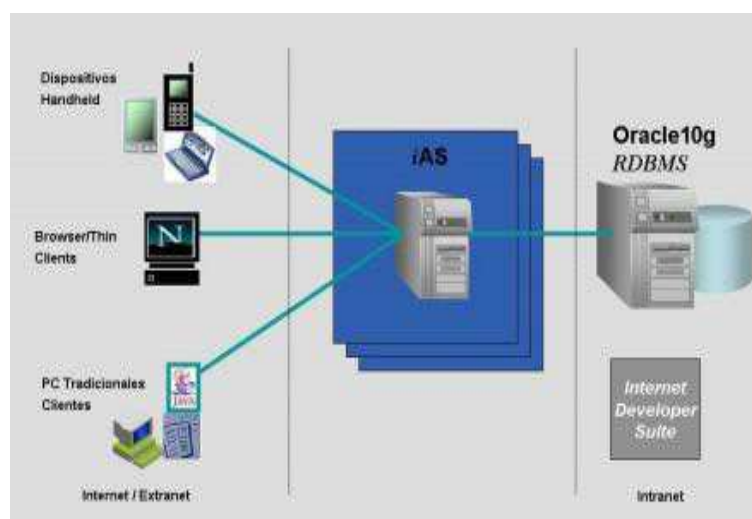


Figura 29 Plataforma de Aplicación de Soluciones

4.3.7. DESARROLLO DEL ETL

4.3.7.1. EXTRACCIÓN DESDE SISTEMAS FUENTE

El Microsoft SQL Server 2005 Integration Services (SSIS) se encargara de realizar la carga de los datos de las RBD ORACLE (área de operaciones), SQL Server (área de marketing) y MySQL.

En la figura 30 mostramos como en el examinador podemos hacer ya uso de la data relacionando las tablas

Dimension				Jerarquía	Operador	Expresión de filtro
<Seleccionar dimensión>						
Customer Geography						
All Customers						
				Product Line		
				Mountain	Road	Accessory
Calendar Year	Calendar Semester	Calendar Quarter	Calendar Month	Internet Sales-Sales Amount	Internet Sales-Sales Amount	Internet Sales-Sales Amount
CY 2001				585.973,27 €	2.680.400,39 €	
CY 2002				1.562.456,76 €	4.967.886,77 €	
CY 2003	H1 CY 2003			1.407.101,03 €	1.630.400,33 €	
	H2 CY 2003	Q3 CY 2003		1.132.099,15 €	1.047.485,08 €	100.387,76 €
		Q4 CY 2003	December 2003	660.835,62 €	568.230,69 €	55.974,28 €
			November 2003	484.649,88 €	396.558,11 €	45.304,78 €
			October 2003	432.877,79 €	353.994,17 €	44.544,14 €
			Total	1.578.363,29 €	1.318.782,97 €	145.823,20 €
	Total	Total	Total	2.710.462,44 €	2.366.268,05 €	246.210,96 €
CY 2004				3.985.190,03 €	2.979.153,04 €	357.842,34 €
Grand Total				10.251.183,52 €	14.624.108,58 €	604.053,30 €

Figura 30 Ejemplo de Consultas

Fuente: Elaboración propia

En la figura 31 observamos como los usuarios pueden generar sus reportes utilizando el Excel , solo tiene que manejar las tablas y seleccionar los datos que deseen , y aquí le mostramos un ejemplo de ello

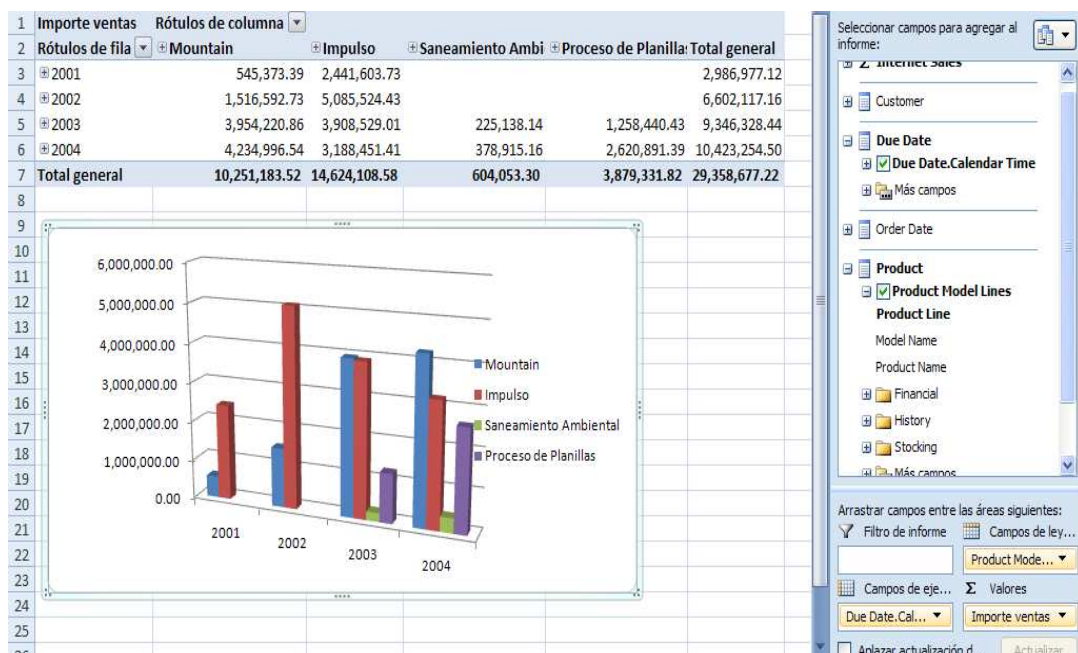


Figura 31 Generación de Reportes

Fuente: Elaboración propia

4.3.7.2. HERRAMIENTAS ETL

La principal herramienta ETL a utilizar en el desarrollo de la aplicación es el Microsoft SQL Server 2005 Integration Services (SSIS), ya que con esta aplicación se realiza la extracción, transformación y carga de los datos.

4.3.7.3. FLUJO DEL DESARROLLO ETL

El flujo para el desarrollo del ETL es:

Extracción: El SSIS extrae la data necesaria de las bases de datos SQL Server, ORACLE y MySQL.

Transformación: Preparación de la data para migrarlo al cubo del DataMart.

Load (carga): Volcado de la información hacia las tablas q conforman el cubo tanto a las tablas de dimensiones como a las Tablas de hechos.

CAPITULO V. CONCLUSIONES Y RECOMENDACIONES

5.1 CONCLUSIONES

Mediante procesos de extracción, transformación y carga de data histórica, de forma automática se obtuvo un repositorio (DataMart) que permitirá la explotación eficiente de la información.

El DataMart permitirá apoyar al área de Logística en la toma de decisiones, a través de la entrega oportuna y relevante de información.

La implementación de un DataMart reduce el tiempo en la elaboración de los reportes tanto al área de logística como al área de Informática, no se necesita tener demasiado conocimiento de los datos almacenados.

El DataMart se constituyó en una herramienta que nos ayuda a la elaboración de reportes y con ello a la toma de decisiones sobre el área de Logística de la empresa T-Impulso.

5.2 RECOMENDACIONES

Si no se tiene manejo de la herramienta SQL Server Integration Services y se quiere reducir costos de implementación, se recomienda utilizar la herramienta PENTAHO ya que es la herramienta de código libre que obtuvo mayor puntaje en el benchmarking.

REFERENCIAS BIBLIOGRÁFICAS

- [1] Conceptos y tecnología de inteligencia de negocio, Agosto 2005, de <http://www.bi-spain.com>
- [2] Historia de la Inteligencia de Negocios, domingo, 21 de junio de 2009, de <http://www.businessintelligence.info/definiciones/historia-business-intelligence.html>.
- [3] [Ing. Bernaben Ricardo Dario, Córdoba, Argentina –17 de Enero de 2009] Data Warehousing: Investigación y Sistematización de conceptos. Metodología para la construcción de un Data Warehouse.
- [4] Inteligencia de Negocios http://www.sinnexus.com/business_intelligence
- [5] [Luis Mendez del Río Octubre 2006] Más allá del Business Intelligence, Unigraf. 181p.
- [6] [NimaRamos 2009] JonathanNimaRamos[Soluciones OLAP con Microsoft SQL Server AnalysisServices](http://www.sinnexus.com/business_intelligence).
- [7] Productos de Oracle para Inteligencia de Negocios, Julio 2008, de <http://www.oracle.com/technology/products/bi/index.html>
- [8] [RaynerHuamantumba 2007]: Manual para el diseño y desarrollo de DataMart.
- [9] [Rios 2011] Angel-Rios; Paper sobre la Inteligencia de Negocios ORACLE.
- [10] [Vincent Rainardi 2008] Building a Data Warehouse with Examples in SQL Server.
- [11] [Vincent Rainardi 2008] Building a Data Warehouse with Examples in SQL Server.
- [12] [Herrera Osorio, Edwar Javier 2011] Diseñando un modulo de inteligencia de negocios con UML
- [13] [Ing. Bernaben Ricardo Dario, Córdoba, Argentina –17 de Enero de 2009]DataWarehousing: Investigación y Sistematización de conceptos. Metodología para la construcción de un Data Warehouse.
- [14] [Booch 199] Patrones de diseño, E. Gamma et al. Editorial Addison-Wesley.
- [15] [Kimball, Ralph] “The Data Warehouse Toolkit series”. Editorial John Wile&Sons.1996-2004.
- [16] [W. H. Inmon]. “Building the Data Warehouse”. Editorial Wiley, TerceraEdición.2002.
- [17] [RaynerHuamantumba 2007]: Manual para el diseño y desarrollo de DataMart.
- [18] [Angel-Rios 2011]; Paper sobre la Inteligencia de Negocios ORACLE

- [19] [Marta Zorilla 2008]: Introducción al Business Intelligence.
- [20] [Luis Mendez del Río Octubre 2006] Más allá del Business Intelligence, 181p.
- [21] [Peña Ayala, Alejandro, Diciembre 2005]: Inteligencia de negocios: Una propuesta para su desarrollo en las Organizaciones, Editorial APA.
- [22] [Luis Cano, Josep 2007] Business Intelligence: Competir con Información
- [23] [Vitt, Elisabeth 2010] Business Intelligence: Técnicas de análisis para la toma de decisiones estratégicas.
- [24] [Adir Even] Managing Metadata in Data Warehouses: Errores y posibilidades.
- [25] [Cano Giner, J. L. 2007]. Business Intelligence: competir con información. Madrid: Fundación Cultural Banesto.
- [26] [Larissa T. Moss, ShakuAtre, 2003]. Business Intelligence Roadmap.
- [27] A brief overview of what the Monte-Carlo method is and does.
<http://www.physics.gla.ac.uk/~donnelly/files/montecarlo/>
- [28] Pagina oficial de Pentaho <http://www.pentaho.com/>
- [29] Claudia Imhoff, Nicholas Galletta, Jonathan G. Geiger. Mastering Data Warehouse Design - Relational and Dimensional Techniques. Indianapolis: Wiley Publishing, Inc., 2003. ISBN: 0 - 471 - 32421 -3